

Der Einfluss von Erwartungen an soziale Interaktionspartner auf Kooperations- und Bestrafungsverhalten

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Laura Mieth
aus Köln

Düsseldorf, Juli 2016

Aus dem Institut für Experimentelle Psychologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

Referent: PD Dr. Raoul Bell

Korreferent: Prof. Dr. Axel Buchner

Tag der mündlichen Prüfung:

Inhaltsverzeichnis

| | |
|--------------------------------|-----|
| Zusammenfassung..... | 4 |
| Abstract..... | 6 |
| Einleitung..... | 7 |
| Experiment 1 | 16 |
| Experiment 2 | 19 |
| Experiment 3 | 23 |
| Experiment 4 | 27 |
| Experiment 5 | 30 |
| Experiment 6 | 34 |
| Allgemeine Diskussion..... | 39 |
| Literatur | 44 |
| Einzelarbeiten..... | 48 |
| Erklärung an Eides Statt | 113 |

Zusammenfassung

Bestimmte Gesichtsm Merkmale können die Erwartung auslösen, dass das Gegenüber eher kooperatives als unkooperatives Verhalten zeigen wird. Die vorliegende Dissertation untersucht, welchen Einfluss diese Erwartungen an Interaktionspartner in einem sozialen Dilemmaspiel auf das Kooperations- und Bestrafungsverhalten der Probanden haben. In sechs Experimenten wurden die Merkmale Sympathie des Gesichts (sympathisch vs. unsympathisch), Gesichtsausdruck (lächelnd vs. neutral) und Geschlecht (weiblich vs. männlich) der Spielpartner sowie das Kooperationsverhalten (Kooperation vs. Defektion) dieser Partner in einem sozialen Dilemmaspiel manipuliert. Die Probanden spielten ein Gefangenendilemmaspiel mit ihren Spielpartnern und konnten diese anschließend für ihr kooperatives oder unkooperatives Verhalten bestrafen. Aufgrund von Stereotypen erwartet man von sympathischen, lächelnden oder weiblichen Spielpartnern eher Kooperation als von unsympathischen, neutral schauenden oder männlichen. Wir interessierten uns dafür, ob erwartungsinkongruentes Verhalten der Spielpartner in einer zweiten Spielphase durch die Probanden mehr bestraft wird als erwartungskongruentes Verhalten. Wenn man die Bestrafung über die Durchgänge aggregiert, zeigt sich eine erhöhte Bestrafung für sympathische und lächelnde im Vergleich zu unsympathischen und neutral schauenden Defektierenden sowie für weibliche im Vergleich zu männlichen Defektierenden. Mit Hilfe von multinomialen Verarbeitungsbaummodellen wurden die Effekte von Kooperations- und Bestrafungsverhalten der Probanden den Spielpartnern gegenüber getrennt voneinander betrachtet. Wie erwartet kooperierten die Probanden mehr mit lächelnden, sympathischen oder weiblichen Spielpartnern (Experimente 1–4). Ebenfalls kooperierten weibliche Probanden mehr mit weiblichen als mit männlichen Spielpartnern, während männliche Probanden keinen Unterschied zwischen weiblichen und männlichen Spielpartnern machten (Experimente 5 und 6). Es zeigten sich keine direkten, negativen Effekte für sympathische, lächelnde oder weibliche Spielpartner in Form von wahrscheinlicherer oder härterer Bestrafung. Die Spielpartner wurden jedoch häufiger moralistisch nach einseitiger Defektion als heuchlerisch nach gegenseitiger Defektion bestraft. Der Effekt von Gesichtsmarkmalen auf die aggregierte Bestrafung lässt sich also allein damit erklären, dass Personen eher mit sympathischen, lächelnden und weiblichen Partnern kooperieren, so dass mehr Gelegenheit für moralistische Bestrafung besteht. Erhöhte Bestrafung von erwartungsinkongruentem im Vergleich zu

erwartungskongruentem Verhalten kann demnach durch indirekte statt durch direkte Effekte erklärt werden.

Abstract

Facial features can lead people to expect cooperation rather than defection from a partner. In the present studies, we investigated the effect of these appearance-based expectations on the participants' cooperation and punishment behavior in a social dilemma game. In six experiments likability (likable vs. unlikable), expression (smiling vs. neutral), and gender (female vs. male) of the partners' faces were manipulated, as well as the partners' behaviors (cooperation vs. defection). Participants played a prisoner's dilemma game in which they could punish their partners for cooperating or defecting. We assumed that participants would expect cooperation from likable and smiling and female faces to a higher degree than from unlikable, neutral, and male faces. We were primarily interested in whether the appearance-incongruent behaviors would be punished disproportionately more. When the total amount of punishment was aggregated over trials, likable, smiling and female partners were indeed punished more for defection than unlikable, non-smiling and male partners. However, it is necessary to disentangle the effects of cooperation and punishment with a multinomial processing tree model to interpret this finding. These analyses showed that participants cooperated more with likable, smiling and female partners (Experiments 1–4). Moreover female participants cooperated more with female than with male partners, while male participants did not treat female and male partners differently (Experiments 5 and 6). However, likable, smiling or female partners did not receive more or harsher punishment than unlikable, neutral or male faces. The moralistic punishment of partners after unilateral defection was higher than the hypocritical punishment after bilateral defection. Thus, people sometimes cooperate more with others based their facial features as heuristic cues, which then gives people more opportunity for moralistic punishment after the unilateral defection of the other person. In conclusion, indirect rather than direct effects lead to the fact that apparently expectancy-incongruent behavior is punished more than expectancy-congruent behavior.

Einleitung

Für das Überleben und die Weiterentwicklung von menschlichen Gesellschaften ist Kooperation von entscheidender Bedeutung. Dabei kann Kooperation durch verschiedene Mechanismen erklärt werden (Nowak, 2006): Verwandtenselektion begünstigt Kooperation zwischen verwandten Individuen, während direkte Reziprozität Kooperation zwischen unverwandten Individuen fördert. Bestrafung von unkooperativem Verhalten kann dabei dafür sorgen, direkte Reziprozität zu begünstigen, weil sie hilft, eine Kooperationsnorm zu verdeutlichen und durchzusetzen. Auch die Theorie der starken Reziprozität (engl. *strong reciprocity*, Gintis, 2000; Gintis, Bowles, Boyd, & Fehr, 2003) besagt, dass Menschen dafür prädisponiert sind, mit anderen zu kooperieren und diejenigen Individuen, die die soziale Kooperationsnorm verletzen, zu bestrafen – selbst wenn diese Bestrafung mit eigenen Kosten verbunden sein mag (engl. *costly punishment*). In vielen Gesellschaften bestrafen Menschen andere für unkooperatives Verhalten, selbst wenn das mit Kosten für das bestrafende Individuum einhergeht (Henrich et al., 2006), da Bestrafung zur Aufrechterhaltung kooperativen Verhaltens beiträgt. Dabei profitieren diese Individuen jedoch nicht unbedingt immer direkt von ihrem Bestrafungsverhalten (Gintis, 2000; Gintis et al., 2003).

Tatsächlich wissen wir, wenn uns eine unbekannte Person zum ersten Mal begegnet, meistens nicht, ob diese daran interessiert ist, sich direkt reziprok zu verhalten. Wir müssen jedoch oft innerhalb kurzer Zeit eine Entscheidung über unser eigenes Verhalten dieser Person gegenüber treffen. Dabei werden verschiedene Informationen in die Formung eines ersten Eindrucks einbezogen (Todorov, 2008). Die Bewertung der Vertrauenswürdigkeit eines Individuums ist dabei nicht nur recht stabil über die Zeit (Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006), sondern findet über verschiedene Personen eine hohe Übereinstimmung (Engell, Haxby, & Todorov, 2007). Die Wahrscheinlichkeit, die Kooperationsneigung eines Partners über einen schnellen, ersten Eindruck korrekt einzuschätzen, liegt allerdings nur knapp über dem Zufallsniveau (Bonnenfon, Hopfensitz, & De Neys, 2013; Fetchenhauer, Groothuis, & Pradel, 2010). Verlässt man sich ausschließlich auf solche ersten Eindrücke, resultiert dies häufig in Fehleinschätzungen, da meistens nur ungenügende Informationen zur Verfügung stehen. Dadurch ergibt sich vor allem bei ersten Begegnungen keine tragfähige Entscheidungsgrundlage, um beurteilen zu können, ob das Gegenüber kooperativ agieren wird. In diesen Fällen basieren Entscheidungen für oder gegen Kooperation auf Heuristiken

oder Stereotypen, die anhand von Aussehen, Mimik und Gestik des Gegenübers schnell und automatisch aktiviert werden (Allport, 1954; Ashmore & Del Boca, 1979). Die Schlussfolgerungen aus Heuristiken und Stereotypen sind dabei nicht immer korrekt. Sie bieten jedoch oft einfache Handlungsempfehlungen. Sicher ist, dass gewisse Merkmale wie zum Beispiel die Attraktivität, der Gesichtsausdruck oder auch das Geschlecht des Interaktionspartners kooperationsbezogene Erwartungen erzeugen können (Eckel & Wilson, 2003; Wilson & Eckel, 2006). Insgesamt werden Menschen mit positiven äußeren Merkmalen oft auch weitere positive Attribute zugeschrieben und es entsteht eine positive Erwartungshaltung bezüglich des Verhaltens dieser Personen (Dion, Berscheid, & Walster, 1972). Wenn das Verhalten einer Person dann von dieser Erwartung abweicht, könnte dies besonders stark bestraft werden. In den vorliegenden sechs Experimenten untersuchten wir, welchen Einfluss das Aussehen eines Menschen, konkret die Sympathie des Gesichts, der Gesichtsausdruck und das Geschlecht, auf Kooperation und Bestrafung dieser Personen in einem Gefangenendilemma hat. In einem Gefangenendilemma können beide Spielpartner entscheiden, entweder zu kooperieren oder zu defektieren. Kooperation bedeutet dabei zur Zusammenarbeit bereit zu sein, während Defektion bedeutet die Zusammenarbeit zu verweigern. Gegenseitige Kooperation führt im Gefangenendilemma zu einem höheren Gewinn als gegenseitige Defektion. Jedoch erzielt man bei einseitiger Defektion den höchsten und bei einseitiger Kooperation den niedrigsten Gewinn. Das Dilemma ergibt sich demnach daraus, dass jeder Partner seinen Gewinn durch Defektion unabhängig vom Verhalten des Partners maximieren kann, während sich jedoch für beide Spielpartner eine bessere Auszahlung ergibt, wenn beide kooperieren als wenn sich beide egoistisch verhalten. Wenn die Defektion eines Gegenübers, von dem kooperatives Verhalten erwartet wurde, als besonders schockierend und enttäuschend erlebt wird, könnte diese Person daraufhin eine vergleichsweise stärkere Bestrafung erfahren als eine defektierende Person, von der aufgrund ihrer äußeren Merkmale kein kooperatives Verhalten erwartet wurde. Ob erwartungsinkongruente Defektion besonders stark bestraft wird, überprüften wir in den vorliegenden Experimenten, in denen den Spielpartnern im Anschluss an das Gefangenendilemma eine Bestrafungsoptionen zur Verfügung stand. Einerseits hätte Bestrafung in diesem Fall die Funktion, eine allgemeine Kooperationsnorm durchzusetzen. Andererseits könnte starke Bestrafung erwartungsinkongruenter Defektion für flexible Merkmale (wie den Gesichtsausdruck) dazu dienen, die Reliabilität der Vorhersage von Kooperation zu begünstigen, während bei inflexiblen Merk-

malen (wie dem Geschlecht) entsprechende Bestrafung dafür sorgen könnte, dass stereotypkonformes Verhalten gefördert und aufrechterhalten wird.

In der Literatur finden sich inkonsistente Befunde zu der Frage, welchen Einfluss Erwartungsverletzungen auf Bestrafung haben. Bestätigende Evidenz dafür, dass erwartungsinkongruentes Verhalten stärker bestraft wird, findet sich beispielsweise in einer Studie zum Vertrauensspiel (Wilson & Eckel, 2006). Beim Vertrauensspiel bekommt Spieler 1, der Treugeber, einen gewissen Geldbetrag. Diesen Geldbetrag kann er entweder komplett behalten oder einen Teil davon an Spieler 2, den Treuhänder, senden. Der gesendete Betrag wird verdreifacht und der Treuhänder kann nun entscheiden wie viel von diesem verdreifachten Betrag er an den Treugeber zurücksenden möchte. Attraktive Treuhänder erhielten mehr Geld als unattraktive (Wilson & Eckel, 2006). Jedoch erhielten attraktive Treugeber weniger Geld zurück als unattraktive, möglicherweise weil sie die Erwartungen, dass sie mehr Geld senden würden, nicht erfüllt hatten. Eine mögliche Erklärung ist, dass Treuhänder von attraktiven Treugebern mehr Geld erwarten als von unattraktiven Treugebern und daher attraktive Treugeber mehr bestrafen, wenn diese die Erwartungen nicht erfüllen. Diese Ergebnisse werden daher im Sinne einer „Schönheitsstrafe“ (engl. *beauty penalty*) interpretiert. Die Entscheidung, von Treuhändern Geld einzubehalten, muss jedoch nicht unbedingt als Bestrafung der Treugeber interpretiert werden, sondern kann auch durch die Tendenz zur eigenen Gewinnmaximierung motiviert sein, da sie direkt davon profitieren. Die Interpretation dieses Verhaltens im Sinne einer moralistischen Bestrafung ist daher problematisch.

Das Ultimatumspiel stellt auf den ersten Blick ein besseres Beispiel für moralistische Bestrafung dar, da Bestrafung für den Bestrafenden mit Kosten verbunden ist, was einer Gewinnmaximierung entgegensteht. Es finden sich dennoch Befunde, die den im vorangegangenen Absatz dargestellten ähneln (Solnick & Schweitzer, 1999). Im Ultimatumspiel bekommt Spieler 1, der Vorschlagende, einen bestimmten Geldbetrag, den er zwischen sich selbst und Spieler 2, dem Antwortenden, aufteilen soll. Der Antwortende akzeptiert diesen Vorschlag oder lehnt ihn ab. Wenn der Teilungsvorschlag angenommen wird, wird das Geld entsprechend aufgeteilt. Bei Ablehnung erhält keiner der beiden Spieler Geld. Ein rationales Verhalten des Antwortenden unter rein ökonomischer Perspektive wäre, jeden Betrag, der angeboten wird, zu akzeptieren, weil jeder Betrag einen Gewinn darstellt. Unfaire Angebote werden jedoch oft zurückgewiesen, was als Form moralistischer Bestra-

fung angesehen werden kann. Solnick und Schweitzer (1999) zeigten, dass attraktiven Antwortenden höhere Angebote gemacht wurden als unattraktiven Antwortenden und dass von attraktiven Vorschlagenden höhere Beträge erwartet wurden als von unattraktiven Vorschlagenden. Schließlich waren jedoch die Ablehnraten für attraktive Vorschlagende höher als für unattraktive Vorschlagende. Weiterhin wurde von Frauen erwartet, dass sie als Vorschlagende mehr Geld anbieten als Männer. Außerdem waren die Ablehnraten für weibliche Vorschlagende höher als für männliche Vorschlagende.

Positive äußere Merkmale scheinen demnach tatsächlich nicht ausschließlich mit positiven Konsequenzen einherzugehen. Vielmehr erzeugen positive Gesichtsmerkmale Erwartungen, welche aber im Falle von erwartungsinkongruentem Verhalten zu stärkerer Bestrafung führen können. Jedoch bleibt in Bezug auf die Ergebnisse von Solnick und Schweitzer (1999) unklar, ob diese unstrittig als moralistische Bestrafung interpretiert werden können, da nicht die Antwort auf unfaire Angebote von Vorschlagenden gemessen wurden, sondern die Antwortenden vor dem eigentlichen Angebot gebeten wurden, ein minimales Angebot festzulegen, welches gerade noch akzeptiert werden würde. Dieses Maß spiegelt wohlmöglich eher eine Erwartung an das zukünftige Verhalten der Vorschlagenden wider als die tatsächliche Motivation, unfaire Angebote zu bestrafen. Diese Befunde können demnach nicht ohne Weiteres generalisiert werden. Sie sollten vielmehr mit einem Paradigma überprüft werden, welches Bestrafungsverhalten in Reaktion auf vorheriges unfaires und egoistisches Verhalten eines Spielpartners direkt misst. Entgegengesetzte Befunde stellen jedoch einen weiteren Grund dar, die Theorie, dass erwartungsinkongruentes Verhalten zu stärkerer Bestrafung führt als erwartungskongruentes Verhalten, ein weiteres Mal zu überprüfen. Im Gegensatz zu den Befunden von Solnick und Schweitzer (1999) berichteten Ma, Hu, Jiang und Meng (2015), dass Probanden im Ultimatumspiel als Antwortende unfaire Angebote von attraktiven Vorschlagenden weniger wahrscheinlich zurückwiesen als von unattraktiven Vorschlagenden. Mussel und Kollegen (Mussel, Göritz, & Hewig, 2013; Mussel, Hewig, Allen, Coles, & Miltner, 2014) fanden, dass unfaire Angebote von lächelnden Vorschlagenden eher akzeptiert wurden als unfaire Angebote von neutral oder wütend schauenden Vorschlagenden. Dies wurde wiederum dahingehend interpretiert, dass Lächeln die negativen Gefühle dem Defektierenden gegenüber abmildern und dadurch für mildere Bestrafung sorgen könnte. Vor dem Hintergrund dieser inkonsistenten Befundlage erscheint weitere Forschung notwendig.

In den vorliegenden Experimenten verwendeten wir eine direkte Bestrafungsoption zur Untersuchung der zugrunde liegenden Mechanismen und des Einflusses von Erwartungen auf Kooperations- und Bestrafungsverhalten in sozialen Interaktionen. Die Probanden spielten in allen nachfolgend beschriebenen Experimenten in Spielphase 1 mit wechselnden Spielpartnern ein simultanes Gefangenendilemmaspiel. Dieses war angelehnt an ein Spiel, welches bereits oft in Untersuchungen zum Quellengedächtnis genutzt wurde (Bell, Buchner, Kroneisen, & Giang, 2012; Bell, Buchner, & Musch, 2010; Bell, Giang, Mund, & Buchner, 2013; Bell et al., 2016; Giang, Bell, & Buchner, 2012). In den vorliegenden Experimenten erhielten beide Spielpartner einen moderaten Gewinn (10 Cent) bei gegenseitiger Kooperation und keinen Gewinn (0 Cent) bei gegenseitiger Defektion. Wenn einer der beiden Spielpartner kooperierte während der andere defektierte, erhielt der einseitig Defektierende die bestmögliche Auszahlung (20 Cent), während einseitige Kooperation zum schlechtestmöglichen Ergebnis (– 10 Cent) führte.

Direkt im Anschluss daran folgte eine zweite Spielphase, in der Geld investiert werden konnte, um den Spielpartner für sein vorheriges Verhalten zu bestrafen. Dabei konnten die Probanden Beträge zwischen 0 und 9 Cent einsetzen, um ihrem Spielpartner das Zehnfache des Einsatzes von dessen Kontostand abziehen zu lassen. Ein Beispieldurchgang des Spiels ist in Abbildung 2 in Mith, Bell und Buchner (2016) dargestellt.

Wir nutzten ein multinomiales Verarbeitungsbaummodell, um Effekte der Gesichtsmerkmale und des Verhaltens der Spielpartner auf das Kooperations- und Bestrafungsverhalten der Probanden zu analysieren und Kooperations- und Bestrafungsbereitschaft voneinander zu trennen. Ursprünglich wurden solche Modelle genutzt, um Schlussfolgerung zu genotypischen Informationen aus beobachtbaren phänotypischen Kategoriehäufigkeiten zu ziehen (Bernstein, 1925). Heutzutage wird mit Hilfe multinomialer Modelle in der Kognitionspsychologie die Wahrscheinlichkeit des Auftretens von latenten kognitiven Zuständen und Prozessen aus beobachtbarem Verhalten geschätzt (Bayen, Murnane, & Erdfelder, 1996; Bröder & Meiser, 2007; Erdfelder et al., 2009; Klauer, Stahl, & Erdfelder, 2007; Riefer & Batchelder, 1988).

Das hier verwendete multinomiale Modell (Abbildung 1) hat fünf Parameter und sechs unabhängige Datenkategorien. Auf der linken Seite werden durch abgerundete Rechtecke die (defektierenden oder kooperierenden) Spielpartner repräsentiert, denen die Probanden im Experiment

begegneten. Parameter C bildet die Wahrscheinlichkeit ab, mit der ein Proband in Spielphase 1 mit einem Spielpartner kooperiert. Die komplementäre Wahrscheinlichkeit $1 - C$ steht dafür, dass der Proband in Spielphase 1 defektiert. Zum Zeitpunkt der Entscheidung des Probanden ist die Entscheidung des Spielpartners noch nicht bekannt. Daher trifft der Proband in Spielphase 1 seine Entscheidung unabhängig von der Kooperationsentscheidung des Spielpartners.

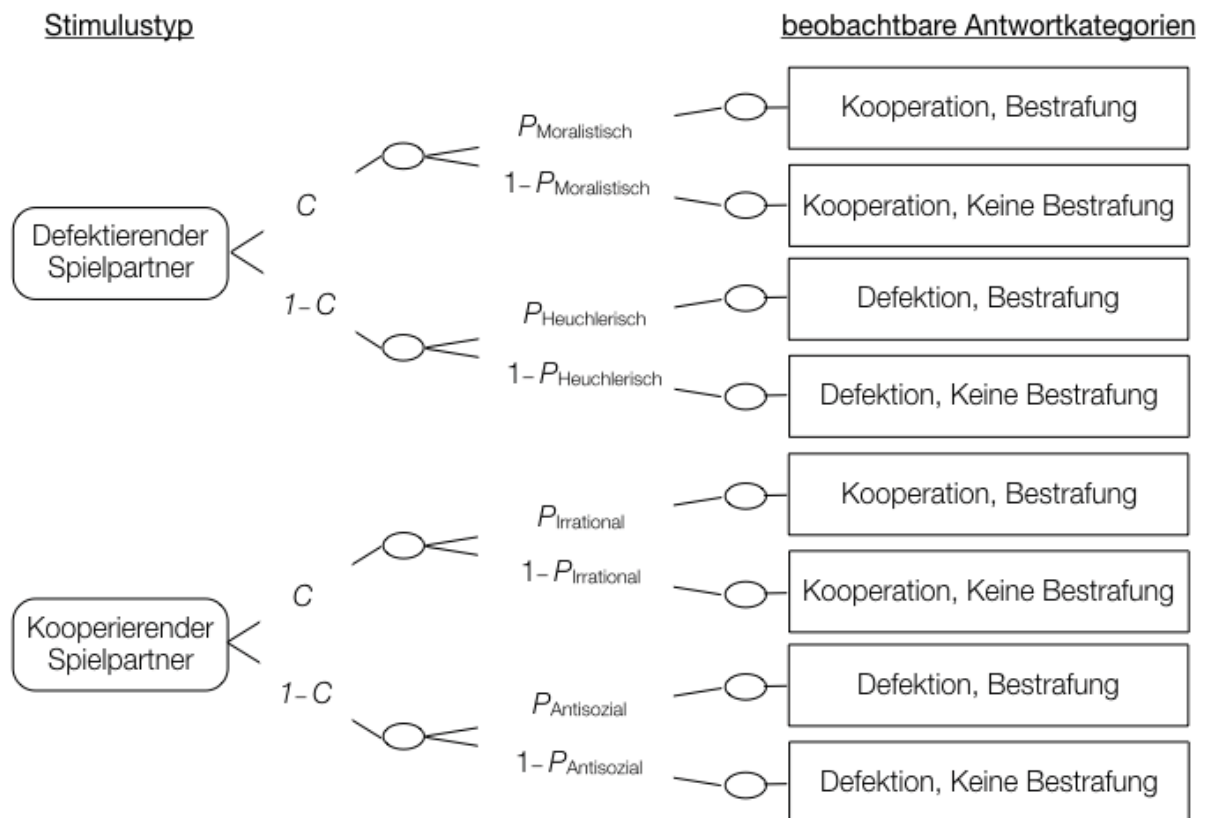


Abbildung 1. Multinomiales Modell zu Kooperation und Bestrafung der Spielpartner durch die Probanden. Die abgerundeten Rechtecke auf der linken Seite repräsentieren die kooperierenden oder defektierenden Spielpartner. Die Rechtecke auf der rechten Seite repräsentieren die beobachtbaren Antwortkategorien der Probanden im Spiel. Parameter C steht für die Wahrscheinlichkeit, dass ein Proband kooperiert. Die Parameter P stehen für die konditionale Wahrscheinlichkeit mit der ein Proband eine bestimmte Art der Bestrafung (moralistisch, heuchlerisch, irrational und antisozial) zeigt.

Das Modell unterscheidet zwischen vier verschiedene Bestrafungsarten, so dass sich vier Parameter ergeben, die das Bestrafungsverhalten der Probanden in Spielphase 2 abhängig vom Verhalten der Spielpartner (Defektion oder Kooperation) und von der eigenen Kooperationsbereitschaft des Probanden (Defektion oder Kooperation) beschreiben. Die P Parameter stehen dabei für konditionalen Wahrscheinlichkeiten, dass der Proband entscheidet seinen Spielpartner zu bestrafen. Die komplementären Wahrscheinlichkeiten ($1 - P$) stehen dafür, dass der Proband entscheidet, seinen

Spielpartner nicht zu bestrafen. Die folgenden vier Bestrafungsarten werden durch das Modell unterschieden: Moralistische Bestrafung (im Modell als Parameter $P_{\text{Moralistisch}}$ bezeichnet) wird definiert als Bestrafung unilateraler Defektion. Diese Bestrafung kann auf kooperatives Verhalten des Probanden in Spielphase 1 folgen, falls der Spielpartner defektiert hat. Moralistische Bestrafung wird häufig beobachtet und kann die Funktion haben, die Kooperationsnorm aufrechtzuerhalten (Carpenter, 2007; Falk, Fehr, & Fischbacher, 2005). Im Falle bilateraler Defektion, bei der also Proband und Spielpartner defektieren, kann der Proband den Spielpartner in Spielphase 2 hypokritisch bzw. heuchlerisch bestrafen (im Modell als $P_{\text{Heuchlerisch}}$ bezeichnet). Diese Art der Bestrafung wird trotz ihres heuchlerischen Charakters ebenfalls oft beobachtet; sie tritt allerdings mit geringerer Wahrscheinlichkeit auf als die moralistische Bestrafung (Carpenter, 2007; Falk et al., 2005). Diese Form von Bestrafung wird als heuchlerisch bezeichnet, weil sie mit üblichen Vorstellungen von Fairness nicht ohne Weiteres in Einklang zu bringen ist und stattdessen durch Boshaftigkeit und dem Wunsch, die Differenz zwischen dem eigenen Gewinn und dem des Gegenübers zu maximieren, motiviert wird (Falk et al., 2005). Gelegentlich bestrafen kooperative Probanden auch kooperative Spielpartner, was aus rationalen Gesichtspunkten schwer zu erklären ist und daher als irrationale Bestrafung bezeichnet wird (im Modell als Parameter $P_{\text{Irrational}}$ bezeichnet). Schließlich zeigt sich antisoziale Bestrafung (im Modell als Parameter $P_{\text{Antisozial}}$ bezeichnet) dann, wenn ein unkooperativer Proband einen kooperativen Spielpartner bestraft. Schätzungsweise 20% allen Bestrafungsverhaltens kann als antisozial klassifiziert werden. Antisoziale Bestrafung wird zuverlässig in verschiedene Gesellschaften gefunden (Herrmann, Thöni, & Gächter, 2008). Da unilaterale Defektion durch kooperative Spielpartner oft geahndet wird, könnte diese Form der Bestrafung eine präventive Vergeltung der antizipierten Bestrafung darstellen. Dadurch, dass die Bestrafungsparameter als konditionale Wahrscheinlichkeiten ausgedrückt werden, können mit Hilfe des Modells Effekte von Gesichtsmerkmalen auf das Bestrafungsverhalten von deren Auswirkung auf das Kooperationsverhalten getrennt werden.

Für die im Folgenden beschriebenen Experimente wurde dieselbe Auswertungsstrategie verwendet. Zunächst wurde der aggregierte Betrag, welcher für Bestrafung ausgegeben wurde (summiert über die Durchgänge und gemittelt über die Probanden) in Abhängigkeit vom jeweils manipulierten Gesichtsmerkmal und dem kooperativen oder unkooperativen Verhalten der Spielpartner analysiert.

Es sollte die Hypothese getestet werden, dass mehr Geld für die Bestrafung von Spielpartnern ausgegeben wird, wenn diese erwartungsin kongruent handeln: Von sympathischen, lächelnden oder weiblichen Spielpartnern wird eher positives, also kooperatives Verhalten erwartet als von unsympathischen, neutral schauenden oder auch männlichen Spielpartnern. Dementsprechend nahmen wir an, dass in die Bestrafung von defektierenden sympathischen, lächelnden und weiblichen Spielpartnern mehr Geld investiert wird, dass also stereotyp- oder erwartungsverletzende Defektion mehr bestraft wird. Für sich gesehen sind diese Analysen bereits interessant. Die Bestrafung in sozialen Interaktionen wird allerdings durch unterschiedliche Faktoren bestimmt: Aussehen und Verhalten der Interaktionspartner haben mutmaßlich ebenso Einfluss auf eine nachfolgende Bestrafung wie das Verhalten der Probanden selbst. Die Bestrafung durch die Probanden hängt somit wohl auch davon ab, ob diese sich in Spielphase 1 selbst kooperativ oder unkooperativ verhalten haben. Auf welche Art und Weise Gesichtsmerkmale einen direkten oder indirekten Einfluss auf das Kooperations- und Bestrafungsverhalten der Probanden haben können, wird in weiteren Analysen erörtert.

Wir testeten die Hypothese, ob die erwartungsin kongruente, unilaterale Defektion eines Spielpartners bei Probanden eine besonders starke emotionale Reaktion auslöst, so dass dieser Spielpartner stärker bestraft wird als ein Spielpartner, der erwartungskongruent defektiert. Ein direkter Effekt von Gesichtsmerkmalen auf Bestrafung könnte also sein, dass sympathische, lächelnde und weibliche Spielpartner wahrscheinlicher moralistisch bestraft werden als unsympathische, neutral schauende und männliche Spielpartner. Insgesamt kommt Bestrafung von Defektion auf Basis des multinomialen Modells mit der Wahrscheinlichkeit $C \cdot P_{\text{Moralistisch}} + (1 - C) \cdot P_{\text{Heuchlerisch}}$ zustande. Angenommen die Wahrscheinlichkeit mit unsympathischen, neutral schauenden oder männlichen Spielpartnern zu kooperieren liegt bei $C = .50$, die Wahrscheinlichkeit, sie moralistisch zu bestrafen, bei $P_{\text{Moralistisch}} = .60$ und die Wahrscheinlichkeit, sie heuchlerisch zu bestrafen, bei $P_{\text{Heuchlerisch}} = .10$. Dann ergibt sich aggregiert, dass unsympathische, neutral schauende oder männliche Spielpartner mit einer Wahrscheinlichkeit von $.50 \cdot .60 + .50 \cdot .10 = .35$ für Defektion bestraft werden. Ein direkter Effekt der Gesichtsmerkmale auf Bestrafung würde bedeuten, dass sympathische, lächelnde oder weibliche Spielpartner mit höherer Wahrscheinlichkeit moralistisch bestraft werden als unsympathische, neutral schauende oder männliche, beispielsweise mit $P_{\text{Moralistisch}} = .90$. Dann ergäbe

sich für sie eine Gesamtwahrscheinlichkeit von $.50 \cdot .90 + .50 \cdot .10 = .50$, für Defektion bestraft zu werden.

Eine stärkere Bestrafung von Spielpartnern, die sich erwartungsinkongruent verhalten und defektieren, muss aber nicht unbedingt auf direkte höher wahrscheinliche Bestrafung zurückzuführen sein. Alternativ könnte dieser Effekt auch durch das Verhalten der Probanden mitbestimmt, also *indirekt* vermittelt sein. Zu dem Grad, zu dem die Probanden ihre Entscheidungen auf Stereotypen oder Heuristiken basieren, sollten sie mit sympathischen, lächelnden und weiblichen Spielpartnern mehr kooperieren als mit unsympathischen, neutral schauenden und männlichen Spielpartnern (Scharlemann, Eckel, Kacelnik, & Wilson, 2001). So ergibt sich die Hypothese, dass mit sympathischen, lächelnden oder weiblichen Spielpartnern mehr kooperiert wird als mit unsympathischen, neutral schauenden oder männlichen Spielpartnern. Dadurch erhöht sich die Wahrscheinlichkeit, unilateraler Defektion ausgeliefert zu sein, und damit die Menge der Gelegenheiten, an denen moralistisch bestraft werden kann. Unabhängig davon ist damit zu rechnen, dass die Probanden Defektion dann mehr bestrafen, wenn sie in Spielphase 1 selbst kooperiert haben, dass sie also mehr moralistische als heuchlerische Bestrafung zeigen. Als Folge der erhöhten Kooperation in Kombination damit, dass wahrscheinlicher moralistisch als heuchlerisch bestraft wird, kann der in Bestrafung investierte, aggregierte Betrag damit indirekt durch positive Erwartungen auslösende Merkmale erhöht werden. Für die Formel $C \cdot P_{\text{Moralistisch}} + (1 - C) \cdot P_{\text{Heuchlerisch}}$ kann dann angenommen werden, dass sympathische, lächelnde oder weibliche Spielpartner und unsympathische, neutral schauende oder männliche Spielpartner mit den gleichen Wahrscheinlichkeiten bestraft werden ($P_{\text{Moralistisch}} = .60$ und $P_{\text{Heuchlerisch}} = .10$), dass sich jedoch für Erstere höhere Kooperationsraten ($C = .80$) ergeben als für Letztere ($C = .50$). Dies würde in der Gesamtanalyse ebenfalls zu einer erhöht wahrscheinlichen Bestrafung von $.80 \cdot .60 + .20 \cdot .10 = .50$ für die Defektion von sympathischen, lächelnden oder weiblichen Spielpartnern im Vergleich zu unsympathischen, neutral schauenden oder männlichen Spielpartnern führen (die mit einer Wahrscheinlichkeit von $.50 \cdot .60 + .50 \cdot .10 = .35$ auftritt).

Das Beispiel verdeutlicht, dass direkte oder indirekte Effekt für eine erhöhte aggregierte Bestrafung erwartungsinkongruenter Defektion sorgen können. Tatsächlich kann auch eine Kombination aus direktem und indirektem Effekt einen aggregierten Effekt erhöhter Bestrafung erwartungs-

inkongruenter Defektion bedingen. Wenn die Probanden mit sympathischen, lächelnden oder weiblichen Spielpartnern eher kooperieren und diese eher moralistisch bestrafen als unsympathische, neutral schauende oder männliche Spielpartner.

Eine weitere zu überprüfende Hypothese ist, dass sympathische, lächelnde oder weibliche Spielpartner härter bestraft werden als unsympathische, neutral schauende oder männliche Spielpartner. Eine härtere Bestrafung bedeutet, dass die Probanden mehr Geld investieren, um eine höhere Bestrafung des Spielpartners zu veranlassen. Hierbei würde es sich ebenfalls um eine *direkte* Auswirkung der Erwartungsverletzung auf die Bestrafung handeln. Die Härte der Bestrafung wurde in einer weiteren Analyse berechnet als der Gesamtgeldbetrag, der in moralistische Bestrafung investiert wurde, geteilt durch die Anzahl der Durchgänge, in denen die Probanden die unilateral defektierenden Spielpartner bestrafen. Es ergibt sich damit notwendigerweise eine Restriktion auf die Substichproben derjenigen Probanden, die in den verglichenen Bedingungen mindestens einmal moralistisch bestrafen. Aus theoretischer Perspektive erscheint die moralistische Bestrafung am interessantesten. Außerdem wird diese Art der Bestrafung häufig genug gezeigt, so dass eine Analyse sinnvoll möglich ist.

Experiment 1

Objektivität und Gleichheit soll rechtem Vorgehen in moralischen Systemen zugrundeliegen. Idealerweise soll also das Aussehen einer Person keinen Einfluss darauf haben, wie normverletzendes Verhalten bestraft wird. Gegenteiliges scheint jedoch oft der Fall zu sein: Einerseits wird potentiell mit sympathisch aussehenden Menschen eher kooperiert, so dass die Motivation Bestrafung einzusetzen steigt, um die Erwidmung dieser Kooperation mit Nachdruck einzufordern. Andererseits verletzt die Defektion durch ein sympathisches Gegenüber Erwartungen, was mit stärkeren Gefühlen von Enttäuschung und Ärger einher geht und somit Bestrafung begünstigen könnte.

Zunächst einmal erschien es interessant, ein vergleichsweise allgemeines Merkmal wie die Sympathie eines Gegenübers, die über dessen Gesicht vermittelt wird, zu betrachten. Die wahrgenommene Sympathie und Vertrauenswürdigkeit des Gesichts korrelieren hoch miteinander (Todorov, 2008; Willis & Todorov, 2006). Mit Hilfe einer Normierungsstudie wurden 20 hoch ($M = 3.79$; $SD = 0.25$) und 20 niedrig sympathische ($M = 1.96$; $SD = 0.27$), weibliche Gesichter aus der

»Center for Vital Longevity«-Datenbank (Minear & Park, 2004) ausgewählt. Jeder Proband interagierte in dem weiter oben beschriebenen Spiel mit zehn sympathisch aussehenden Spielpartnern, die kooperierten, und mit zehn sympathisch aussehenden Spielpartnern, die defektierten. Das Gleiche galt für die unsympathisch aussehenden Spielpartner. Die computergesteuerten Spielpartner waren so programmiert, dass ausschließlich Spielpartner, die zuvor kooperiert hatten, defektierenden Probanden eine randomisierte Strafe zwischen 1 und 10 Cent auferlegten.

Zunächst zeigte sich, dass von 80 Probanden (davon 56 Frauen) 72 Defektion mindestens einmal bestraften, während 37 Probanden ebenfalls Kooperation mindestens einmal bestraften. Für den aggregierten Betrag (aufsummiert über die Durchgänge und gemittelt über die Probanden), der in die Bestrafung der Spielpartner investiert wurde (Tabelle 1), ergab sich ein signifikanter Haupteffekt des Verhaltens der Spielpartner [$F(1,79) = 87.93, p < .001, \eta_p^2 = .53$] dahingehend, dass die Probanden mehr in die Bestrafung defektierender Spielpartner investierten. Der Haupteffekt der Sympathie des Gesichts war auch signifikant [$F(1,79) = 7.55, p = .007, \eta_p^2 = .09$] und reflektierte die höheren investierten Beträge in die Bestrafung sympathisch aussehender verglichen mit unsympathisch aussehenden Spielpartnern. Die Interaktion zwischen Verhalten des Spielpartners und dessen Aussehen stellte sich als ebenfalls signifikant heraus [$F(1,79) = 13.40, p < .001, \eta_p^2 = .14$]. Die Probanden investierten mehr Geld in die Bestrafung sympathischer als unsympathischer defektierender Spielpartner, und sie investierten mehr Geld in die Bestrafung unsympathischer als sympathischer Kooperierender. In Experiment 1 bestätigte sich somit, dass mehr Geld für die Bestrafung von Spielpartnern ausgegeben wurde, wenn ihr Verhalten erwartungskongruent war. Im Folgenden wird nun erläutert, wie neben Aussehen und Verhalten der Spielpartner auch das Kooperationsverhalten der Probanden in Spielphase 1 Einfluss auf das Bestrafungsverhalten der Probanden hatte.

Tabelle 1: Aggregierter Betrag, den die Probanden in Experiment 1 ausgaben, um ihre Spielpartner zu bestrafen (summiert über die Durchgänge und gemittelt über die Probanden) in Cent für die Kombinationen aus Sympathie des Gesichts (sympathisch vs. unsympathisch) und Verhalten des Spielpartners (Defektion vs. Kooperation).

| Sympathie des Gesichts | Verhalten des Spielpartners | | | |
|------------------------|-----------------------------|---------------|-------------|---------------|
| | Defektion | | Kooperation | |
| | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) |
| Sympathisch | 24.26 | (2.33) | 2.65 | (0.79) |
| Unsympathisch | 19.15 | (2.40) | 3.94 | (0.99) |

Für die modellbasierte Analyse wurden zwei Sets des multinomialen Modells benötigt (Abbildung 1), eines für sympathische und eines für unsympathische Spielpartner. Modellpassung und statistische Vergleiche wurden mit Hilfe des Programms multiTree durchgeführt (Moshagen, 2010). Im Basismodell ergab sich eine gute Modellpassung [$G^2(2) = 1.01, p = .604$]. Die Probanden kooperierten eher mit sympathisch als mit unsympathisch aussehenden Spielpartnern [$\Delta G^2(1) = 202.92, p < .001, w = 0.25$]. Alle Parameterschätzer sind in Abbildung 2 abgebildet.

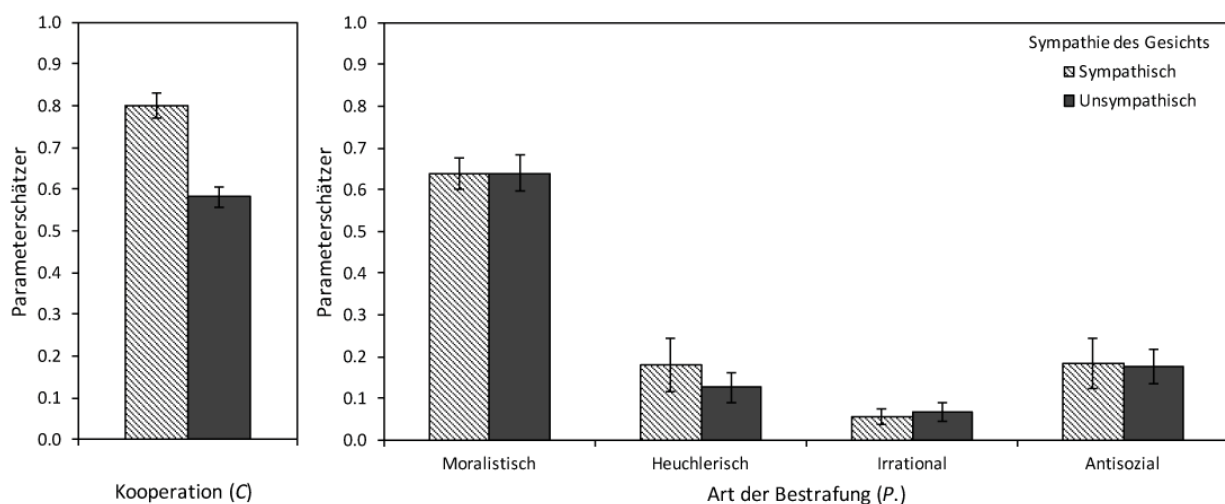


Abbildung 2. Parameterschätzer für Experiment 1. In der linken Grafik sind die Kooperationsparameter (C) abgebildet. Sie repräsentieren die Wahrscheinlichkeiten, dass die Probanden kooperiert haben in Abhängigkeit von der Sympathie des Gesichts der Spielpartner (sympathisch vs. unsympathisch). In der rechten Grafik sind die Bestrafungsparameter (P.) in Abhängigkeit von der Art der Bestrafung (moralistisch, heuchlerisch, irrational und antisozial) und der Sympathie des Gesichts der Spielpartner (sympathisch vs. unsympathisch) abgebildet. Die Fehlerbalken repräsentieren die .95-Konfidenzintervalle.

Es zeigte sich eine höhere Wahrscheinlichkeit für moralistische als für heuchlerische Bestrafung [$\Delta G^2(2) = 334.01, p < .001, w = 0.32$]. Irrationale Bestrafung war weniger wahrscheinlich als antisoziale Bestrafung [$\Delta G^2(2) = 46.16, p < .001, w = 0.12$].

Die Wahrscheinlichkeit moralistischer Bestrafung unterschied sich nicht zwischen sympathischen und unsympathischen Spielpartnern [$\Delta G^2(1) < 0.01, p = .967, w < 0.01$]. Das Gleiche galt für heuchlerische [$\Delta G^2(1) = 2.30, p = .130, w = 0.03$], irrationale [$\Delta G^2(1) = 0.46, p = .495, w = 0.01$] und antisoziale Bestrafung [$\Delta G^2(1) = 0.04, p = .834, w < 0.01$]. Um die Analysen zu komplettieren, wurde schließlich die Härte der moralistischen Bestrafung analysiert. Wir testeten, ob sich härtere moralistische Bestrafung für sympathische im Vergleich zu unsympathischen defektierenden Spielpartnern ergab. Die Härte der moralistischen Bestrafung (deskriptive Statistiken in Mieth, Bell, et al., 2016, Experiment 1, Results) unterschied sich nicht für sympathisch und unsympathisch aussehende Spielpartner [$F(1,65) = 0.42, p = .519, \eta_p^2 < .01$].

Zunächst einmal wurde also auf aggregierter Ebene die erwartungsinkongruente Defektion sympathisch aussehender Spielpartner mehr bestraft als Defektion unsympathisch aussehender Spielpartner. Die Annahme, dass Personen, die sich erwartungsinkongruent verhalten, wahrscheinlicher oder härter moralistisch bestraft werden, konnte jedoch nicht bestätigt werden. Die Analyse der zugrundeliegenden Prozesse zeigte, dass die Probanden mehr mit sympathischen als mit unsympathischen Spielpartnern kooperierten. Der Effekt erhöhter aggregierter Bestrafung erwartungsinkongruenter Defektion im Vergleich zu erwartungskongruenter Defektion war damit indirekt vermittelt über die erhöhte Kooperationsbereitschaft sympathisch aussehenden Spielpartnern gegenüber. Aus dieser erhöhten Kooperationsbereitschaft heraus ergab sich, dass mehr Gelegenheit zu moralistischer Bestrafung bestand, wenn die Erwartung gegenseitiger Kooperation durch das sympathische Gegenüber enttäuscht wurde.

Experiment 2

Der Gesichtsausdruck der Spielpartner wurde in Experiment 2 dahingehend manipuliert, dass lächelnde und neutral schauende Spielpartner präsentiert wurden. Damit diene Experiment 2 als konzeptuelle Replikation von Experiment 1, fügte jedoch einen weiteren Aspekt hinzu. Während die Sympathie eines Gesichts ein vergleichsweise stabiles Merkmal darstellt, ist der Gesichtsausdruck

zum Teil willentlich beeinflussbar und kann somit spezifische Signale senden. Lächeln kann genutzt werden, um Prosozialität und Kooperationsbereitschaft zu signalisieren (Eckel & Wilson, 2003). Die Wahrscheinlichkeit und Härte moralistischer Bestrafung wurde nicht direkt durch die Sympathie des Gesichts des Spielpartners beeinflusst (Experiment 1). Es erscheint jedoch möglich, dass Defektion eines lächelnden Gegenübers eine noch stärkere negative Emotion hervorruft als wenn ein sympathisches Gegenüber defektiert, weil Lächeln kooperative Intentionen signalisiert. In dem Fall könnten die Probanden die einseitige Defektion eines lächelnden Spielpartners mit höherer Wahrscheinlichkeit oder härter bestrafen als die einseitige Defektion eines neutral schauenden Spielpartners (Brown & Moore, 2002). Die entgegengesetzte Hypothese lässt sich aber ebenfalls theoretisch herleiten. Defektierende Spielpartner könnten aufgrund von Lächeln mit Nachsicht behandelt und entsprechend milde bestraft werden, weil Lächeln positive Emotionen im Gegenüber auslösen kann, so dass eine negative Bewertung abgeschwächt wird und Lächeln besänftigend wirkt (Goldenthal, Johnston, & Kraut, 1981; Mackey, 1976).

Den Probanden wurden Bilder weiblicher Gesichter aus der »Center for Vital Longevity«-Datenbank (Minear & Park, 2004) gezeigt, die lächelten oder einen neutralen Gesichtsausdruck hatten. In den 40 Durchgängen des sozialen Interaktionsspiels kooperierten und defektierten jeweils zehn lächelnde und zehn neutral schauende Spielpartner. Auch in Experiment 2 waren die computergesteuerten Spielpartner so programmiert, dass Spielpartner, die zuvor kooperiert hatten, defektierenden Probanden eine randomisierte Strafe auferlegten.

Von den 80 Probanden (davon 58 Frauen) bestrafte 73 Defektion mindestens einmal. Außerdem bestrafte 38 von 80 Probanden Kooperation mindestens einmal. Aggregiert über alle Durchgänge hinweg (Tabelle 2) gaben die Probanden mehr Geld für die Bestrafung defektierender als für die Bestrafung kooperierender Spielpartner aus [$F(1,79) = 59.10, p < .001, \eta_p^2 = .43$]. Der Haupteffekt des Gesichtsausdrucks wurde nicht signifikant [$F(1,79) = 1.87, p = .176, \eta_p^2 = .02$], jedoch zeigte sich eine signifikante Interaktion [$F(1,79) = 12.78, p < .001, \eta_p^2 = .14$] dahingehend, dass mehr Geld investiert wurde, um lächelnde Defektierende im Vergleich zu neutral schauenden Defektierenden zu bestrafen. Die Probanden gaben ebenfalls mehr Geld aus, um neutral schauende Kooperierende im Vergleich zu lächelnden Kooperierenden zu bestrafen. Analog zu Experiment 1 bestätigte sich also, dass für die Bestrafung von erwartungsinkongruenter Defektion mehr Geld

ausgegeben wurde, als für die Bestrafung erwartungskongruenter Defektion. Die zugrundeliegenden Mechanismen wurden mit Hilfe multinomialer Analysen identifiziert.

Tabelle 2: Aggregierter Betrag, den die Probanden in Experiment 2 ausgaben, um ihre Spielpartner zu bestrafen (summiert über die Durchgänge und gemittelt über die Probanden) in Cent für die Kombinationen aus Gesichtsausdruck (lächelnd vs. neutral schauend) und Verhalten des Spielpartners (Defektion vs. Kooperation).

| Gesichtsausdruck | Verhalten des Spielpartners | | | |
|------------------|-----------------------------|---------------|-------------|---------------|
| | Defektion | | Kooperation | |
| | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) |
| Lächelnd | 22.33 | (2.55) | 2.86 | (0.73) |
| Neutral | 18.80 | (2.27) | 4.45 | (0.89) |

In der multinomialen Analyse ergab sich eine gute Passung der Daten zum Modell [$G^2(2) = 0.28$, $p = .879$]. Die Probanden zeigten eine höhere Bereitschaft, mit lächelnden im Vergleich zu neutral schauenden Spielpartnern zu kooperieren [$\Delta G^2(1) = 105.47$, $p < .001$, $w = 0.18$]. Alle Parameter sind in Abbildung 3 abgebildet.

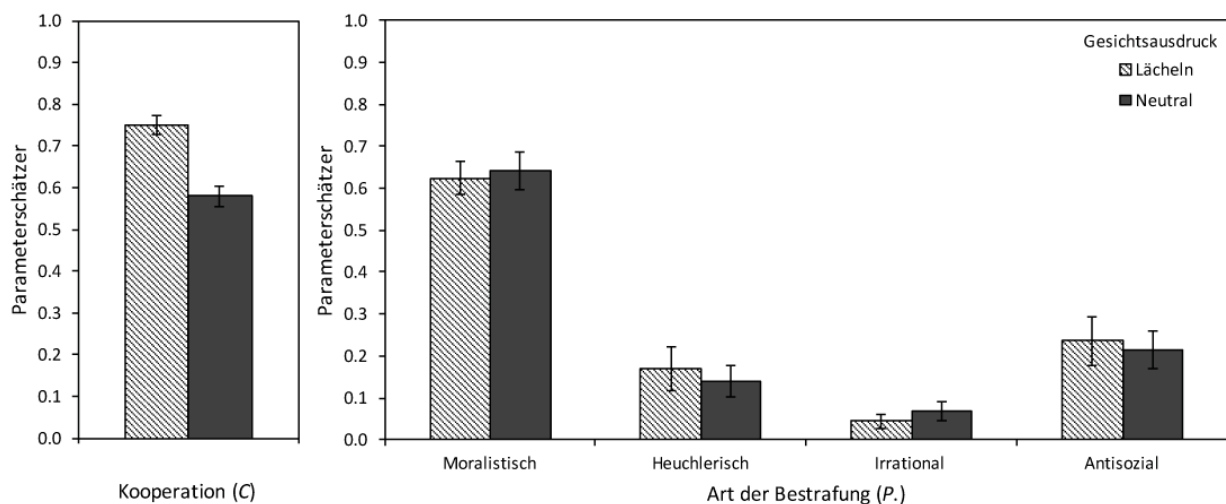


Abbildung 3. Parameterschätzer für Experiment 2. In der linken Grafik sind die Kooperationsparameter (C) abgebildet. Sie repräsentieren die Wahrscheinlichkeiten, dass die Probanden kooperiert haben in Abhängigkeit vom Gesichtsausdruck der Spielpartner (lächelnd vs. neutral). In der rechten Grafik sind die Bestrafungsparameter (P.) in Abhängigkeit von der Art der Bestrafung (moralistisch, heuchlerisch, irrational und antisozial) und des Gesichtsausdrucks der Spielpartner (lächelnd vs. neutral) abgebildet. Die Fehlerbalken repräsentieren die .95-Konfidenzintervalle.

Moralistische Bestrafung wurde signifikant wahrscheinlicher angewandt als heuchlerische Bestrafung [$\Delta G^2(2) = 347.23, p < .001, w = 0.33$]. Weiterhin fand sich, dass irrationale Bestrafung mit signifikant geringerer Wahrscheinlichkeit gezeigt wurde als antisoziale Bestrafung [$\Delta G^2(2) = 93.29, p < .001, w = 0.17$].

Jedoch unterschied sich die Wahrscheinlichkeit für moralistische Bestrafung nicht zwischen lächelnden und neutral schauenden Spielpartnern [$\Delta G^2(1) = 0.31, p = .575, w = 0.01$]. Gleiches galt für heuchlerische [$\Delta G^2(1) = 0.85, p = .358, w = 0.02$], irrationale [$\Delta G^2(1) = 2.29, p = .131, w = 0.03$] und antisoziale Bestrafung [$\Delta G^2(1) = 0.30, p = .582, w = 0.01$]. Schließlich testeten wir, ob sich unterschiedlich harte Bestrafung für lächelnde im Vergleich zu neutral schauenden Defektierenden ergab. Für die Härte moralistischer Bestrafung (deskriptive Statistiken in Mieth, Bell, et al., 2016, Experiment 2, Results) fanden wir, dass lächelnde Defektierende weniger hart moralistisch bestraft wurden als neutral schauende Defektierende [$F(1,66) = 5.77, p = .019, \eta_p^2 = .08$].

Damit wurden die Hauptergebnisse von Experiment 1 auch in Experiment 2 bestätigt. Zunächst wurde erwartungsinkongruentes Verhalten mit erhöhtem Einsatz bestraft. Dieser Effekt verletzter Erwartungen wurde jedoch nicht durch direkte höhere Wahrscheinlichkeit oder härtere moralistische Bestrafung lächelnder Defektierender vermittelt, sondern ergab sich vielmehr dadurch, dass mit lächelnden Spielpartnern mehr kooperiert wurde als mit nicht lächelnden. Außerdem wurde mehr moralistisch als heuchlerisch bestraft, so dass mehr Gelegenheit für moralistische Bestrafung bestand. Bezüglich der Härte moralistischer Bestrafung zeigte sich sogar eine gewisse Milde gegenüber lächelnden verglichen mit neutral schauenden Spielpartnern. Dies passt zu Befunden, die zeigen konnten, dass Lächeln die Bestrafung in sozialen Spielen reduzieren kann (Mussel et al., 2013). Außerdem wird das Erhalten negativen Feedbacks mit einer negativen neuronalen Komponente (engl. *feedback related negativity*) in Zusammenhang gebracht. Diese negative Komponente fällt jedoch nach einem unfairen Angebot eines lächelnden verglichen mit einem neutral schauenden Gegenüber weniger stark aus (Mussel et al., 2014). Somit scheint Lächeln nicht nur Vertrauen und kooperative Intentionen zu signalisieren (Eckel & Wilson, 2003), sondern auch eine besänftigende Funktion zu haben und Aggression abzubauen (Goldenthal et al., 1981; Mackey, 1976). Eine allgemeinere Erklärung wäre, dass ein lächelndes Gegenüber im Betrachter positive Emotionen auslöst, so dass die negative Bewertung einer Defektion des Gegenübers weniger stark ausfällt.

Experiment 3

Experiment 3 stellte eine weitere konzeptuelle Replikation der bisherigen Befunde dar. Darüber hinaus sollte ein weiterer Aspekt untersucht werden: Welche Auswirkung hatte die moralistische Bestrafung der Probanden durch die Spielpartner auf Kooperations- und Bestrafungsverhalten? In den Experimenten 1 und 2 waren die Spielpartner so programmiert, dass defektierende Probanden durch kooperierende Spielpartner moralistisch bestraft wurden. Diese Erfahrung moralistischer Bestrafung wird typischerweise auch bei echten menschlichen Interaktionen gemacht (Carpenter, 2007; Falk et al., 2005). Durch die moralistische Bestrafung der Spielpartner könnte bei den Probanden das Erlernen einer Norm begünstigt worden sein, die einerseits Kooperation seitens der Probanden gefördert haben kann und sie andererseits dazu animiert haben kann, ihre Spielpartner moralistisch zu bestrafen. Es liegt daher nahe anzunehmen, dass Kooperations- und Bestrafungsverhalten der Probanden von der moralistischen Bestrafung durch die Spielpartner beeinflusst wurden. Unklar ist jedoch, wie sich dies auf Effekte der Sympathie des Gesichts oder des Gesichtsausdrucks auf das Kooperations- und Bestrafungsverhalten der Probanden ausgewirkt haben könnte. In Experiment 3 untersuchten wir diesen Einfluss, indem die Spielpartner nicht mehr bestraften und somit nicht zum Erlernen einer Bestrafungsnorm beitrugen. Die Bestrafungsoption stand ausschließlich den Probanden zur Verfügung, während keiner der Spielpartner bestrafte. Abgesehen davon war Experiment 3 identisch zu Experiment 1 – es wurden sympathische und unsympathische Spielpartner präsentiert, die kooperierten oder defektierten.

Im Sinne starker Reziprozität sorgt moralistische Bestrafung für die Aufrechterhaltung von Kooperation in sozialen Gruppen (Fehr & Gächter, 2002; Gintis, 2000; Gintis et al., 2003). Falls die Probanden in den ersten beiden Experimenten nur kooperierten, weil sie befürchteten, für unkooperatives Verhalten durch ihre Spielpartner moralistisch bestraft zu werden, sollte die Kooperationsbereitschaft der Probanden in Experiment 3, in dem keine Bestrafung zu befürchten war, zurückgehen. Wir nahmen jedoch an, dass die Probanden auch ohne das modellhafte Bestrafungsverhalten ihrer Spielpartner weiterhin am wahrscheinlichsten moralistisch bestrafen würden. Außerdem sollten die Probanden auch in Experiment 3 eher mit sympathischen als mit unsympathischen Spielpartnern kooperieren. Und schließlich erwarteten wir einen indirekten Effekt der Sympathie des Gesichts auf moralistische Bestrafung zu finden.

Von den 80 Probanden (davon 56 Frauen) bestraften 64 mindestens einmal eine Defektion, während 26 mindestens einmal eine Kooperation bestraften. Für den aggregierten Betrag, der für die Bestrafung der Spielpartner ausgegeben wurde (Tabelle 3), ergab sich, dass die Probanden mehr in die Bestrafung von defektierenden als von kooperierenden Spielpartnern investierten [$F(1,79) = 26.64, p < .001, \eta_p^2 = .25$]. Der Haupteffekt der Sympathie des Gesichts war nicht signifikant [$F(1,79) = 2.33, p = .131, \eta_p^2 = .03$]. Die Interaktion zwischen Sympathie des Spielpartners und deren Verhalten auf das Bestrafungsverhalten der Probanden war ebenfalls nicht signifikant [$F(1,79) = 2.29, p = .134, \eta_p^2 = .03$]. Tendenziell wurde auch in Experiment 3 mehr Geld in die Bestrafung sympathischer Defektierender als in die Bestrafung unsympathischer Defektierender investiert [$F(1,79) = 3.47, p = .066, \eta_p^2 = .04$]. Wir analysierten mit Hilfe des multinomialen Modells, welchen Einfluss die Entscheidungen der Probanden mit sympathischen und unsympathischen Spielpartnern zu kooperieren hatte.

Tabelle 3: Aggregierter Betrag, den die Probanden in Experiment 3 ausgaben, um ihre Spielpartner zu bestrafen (summiert über die Durchgänge und gemittelt über die Probanden) in Cent für die Kombinationen aus Sympathie des Gesichts (sympathisch vs. unsympathisch) und Verhalten des Spielpartners (Defektion vs. Kooperation).

| | Verhalten des Spielpartners | | | |
|------------------------|-----------------------------|---------------|-------------|---------------|
| | Defektion | | Kooperation | |
| Sympathie des Gesichts | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) |
| Sympathisch | 10.86 | (1.66) | 2.86 | (0.80) |
| Unsympathisch | 9.23 | (1.59) | 3.13 | (0.90) |

In der multinomialen Analyse fand sich eine gute Passung der Daten zum Modell [$G^2(2) = 1.78, p = .411$]. Die Wahrscheinlichkeit, mit der die Probanden in Experiment 3 kooperierten, war im Vergleich zu Experiment 1 geringer, vermutlich weil Probanden sich nicht vor moralistischer Bestrafung durch ihre Spielpartner fürchten mussten. Trotzdem fanden wir kooperatives Verhalten seitens der Probanden: Probanden kooperierten mehr mit sympathischen im Vergleich zu unsympathischen Spielpartnern [$\Delta G^2(1) = 106.00, p < .001, w = 0.18$]. Alle Parameter sind in Abbildung 4 abgebildet.

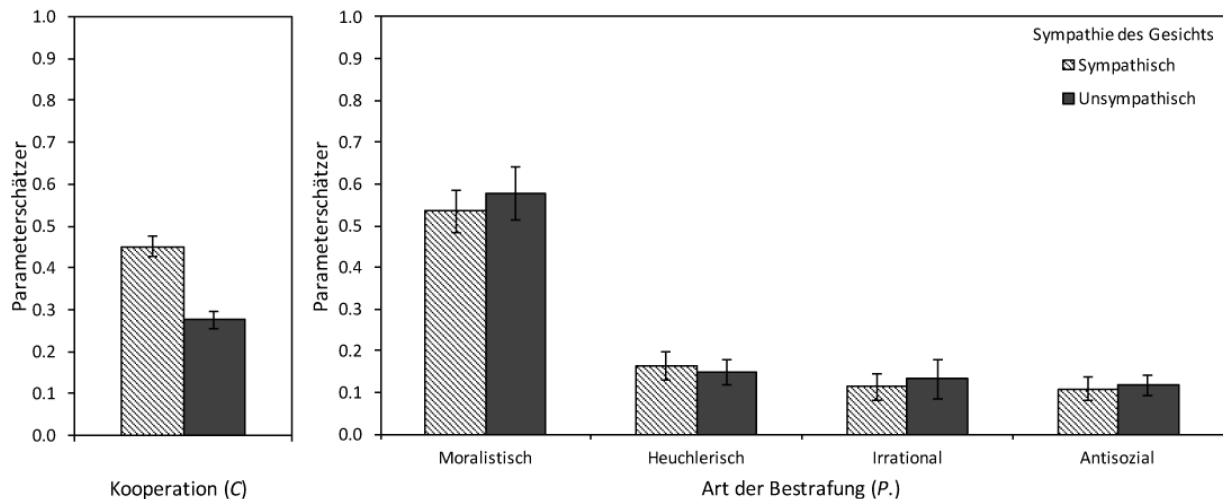


Abbildung 4. Parameterschätzer für Experiment 3. In der linken Grafik sind die Kooperationsparameter (C) abgebildet. Sie repräsentieren die Wahrscheinlichkeiten, dass die Probanden kooperiert haben in Abhängigkeit von der Sympathie des Gesichts der Spielpartner (sympathisch vs. unsympathisch). In der rechten Grafik sind die Bestrafungsparameter (P) in Abhängigkeit von der Art der Bestrafung (moralistisch, heuchlerisch, irrational und antisozial) und der Sympathie des Gesichts der Spielpartner (sympathisch vs. unsympathisch) abgebildet. Die Fehlerbalken repräsentieren die .95-Konfidenzintervalle.

Moralistische Bestrafung trat signifikant wahrscheinlicher auf als heuchlerische Bestrafung [$\Delta G^2(2) = 269.72, p < .001, w = 0.29$]. Weiterhin ergab sich, dass irrationale Bestrafung genauso wahrscheinlich gezeigt wurde wie antisoziale Bestrafung [$\Delta G^2(2) = 0.38, p = .836, w = 0.01$]. Das stellt eine Veränderung im Vergleich zu Experiment 1 dar, in dem antisoziale Bestrafung wahrscheinlicher auftrat als irrationale Bestrafung. Diese Veränderung war zu erwarten gewesen, weil antisoziale Bestrafung zumindest zum Teil durch die präventive Vergeltung der antizipierten moralistischen Bestrafung motiviert wird (Nikiforakis, 2008).

Wie erwartet – und wie in den Experimenten 1 und 2 – hatte Sympathie des Gesichts keinen Einfluss auf die Wahrscheinlichkeit moralistischer Bestrafung [$\Delta G^2(1) = 1.06, p = .304, w = 0.02$]. Das Gleiche galt für heuchlerische [$\Delta G^2(1) = 0.43, p = .511, w = 0.01$], irrationale [$\Delta G^2(1) = 0.45, p = .504, w = 0.01$] und antisoziale Bestrafung [$\Delta G^2(1) = 0.15, p = .696, w = 0.01$]. Und auch für die Härte der Bestrafung (deskriptive Statistiken in Mieth, Bell, et al., 2016, Experiment 3, Results) fand sich erneut, dass sympathische und unsympathische Spielpartner gleich hart bestraft wurden [$F(1,44) = 0.80, p = .375, \eta_p^2 = .02$].

In Experiment 3 investierten die Probanden nur tendenziell mehr Geld in die Bestrafung von erwartungsinkongruenter Defektion. Jedoch zeigten sich weiterhin parallele Ergebnisse zu den beiden vorherigen Experimenten: Die Probanden kooperierten mehr mit sympathischen als mit unsympathischen Spielpartnern. Die Sympathie des Gesichts hatte keinen direkten Effekt auf die Wahrscheinlichkeit oder die Härte moralistischer Bestrafung. Außerdem war die Wahrscheinlichkeit moralistischer Bestrafung im Vergleich zu den anderen Bestrafungsarten erhöht, obwohl die Spielpartner nicht mehr als Vorbild bezüglich dieser Bestrafungsnorm dienten.

Moralistische Bestrafung soll Kooperation fördern. Die Kooperation ging in Experiment 3 verglichen mit den vorherigen Experimenten zurück. Dies war zu erwarten, da die Probanden in den Experimenten 1 und 2 von ihren Spielpartnern moralistisch bestraft wurden, wenn sie unilateral defektierten. In Experiment 3 konnten die Probanden unilateral defektieren, ohne dass sie eine Bestrafung erfuhren. Aus ökonomischer Perspektive sollten die Probanden daher ausschließlich defektieren, um ihren Gewinn zu maximieren. Unsere Daten bestätigen nicht, dass Probanden sich ausschließlich rational verhalten. Die Probanden kooperierten auch in Experiment 3 mit ihren Spielpartnern. Folglich handelten auch die Probanden der vorherigen Experimente mutmaßlich nicht ausschließlich aus Angst vor moralistischer Bestrafung kooperativ, sondern auch weil sie einer Kooperationsnorm folgten. Diese Norm wurde jedoch nicht im Spiel vorgegeben, sondern entsprang einem echten Bedürfnis nach fairen Aktionen und Reaktionen. Die Investition der Probanden in die Bestrafung ihrer Spielpartner legt ebenfalls eine immanente Kooperationsnorm nahe, da die Probanden mit jedem Spielpartner nur einmal interagierten. Es ist nicht rational, Geld für die Bestrafung eines Gegenübers auszugeben, mit dem man nicht mehr interagiert, weil man aus der Bestrafung (und möglichen positiven Konsequenzen) keinen Nutzen ziehen kann. Die Probanden können demnach nicht ausschließlich darauf bedacht gewesen sein, ihren eigenen Gewinn zu maximieren. Es hat vielmehr den Anschein, als hätte die Bestrafung als solche eine belohnende Funktion. Das bestätigt frühere Befunde, dass Menschen gewillt sind, Geld in die Bestrafung unkooperativen oder unfairen Verhaltens zu investieren (Carpenter, 2007; Falk et al., 2005; Fehr & Gächter, 2002).

Experiment 4

Verschiedene Gesichtsmerkmale können unterschiedlich subtil oder salient sein. Das in Experiment 2 untersuchte Merkmal Lächeln ist sogar eines, das von Interaktionspartnern in gewissen Grenzen situativ angepasst werden kann. Die Geschlechtszugehörigkeit des Gegenübers ist hingegen ein Merkmal, welches nicht variabel und gleichzeitig höchst salient ist. Da mit dem Geschlecht häufig sehr präzise Stereotype und Charakteristiken assoziiert werden, wird auch die Geschlechterzugehörigkeit des Gegenübers genutzt, um Informationen über die Kooperationsbereitschaft abzuleiten (Snyder, Tanke, & Berscheid, 1977). Geschlechterstereotype lassen Frauen oft in einem positiven Licht erscheinen. Beide Geschlechter schreiben Frauen seltener eigennützige Eigenschaften wie kompetitiv oder dominant, sondern eher selbstlose Eigenschaften wie hilfsbereit und gütig zu (Eagly & Mladinic, 1989). Solche wohlwollenden Geschlechterstereotype implizieren, dass sich Frauen im Vergleich zu Männern weniger egoistisch sondern eher kooperativ verhalten, auch wenn die aktuelle Forschung zu Unterschieden im Kooperationsverhalten zwischen Männern und Frauen eher gemischte Befunde verzeichnet (Balliet, Li, Macfarlan, & Van Vugt, 2011). In Experiment 4 prüften wir die Hypothese, ob sich diese positiven Erwartungen an Frauen in erhöhter Bestrafung niederschlagen, wenn Frauen erwartungsinkongruentes Verhalten zeigen, also Verhalten, welches nicht der weiblichen Geschlechterrolle entspricht. Eine Frau, die sich egoistisch verhält, müsste dann befürchten, mehr bestraft zu werden als ein Mann, der sich ebenso verhält. Wohlwollende Geschlechterstereotype gegenüber Frauen könnten damit zu negativen Folgen führen, wenn Frauen dadurch Bestrafung und sozialen Druck erfahren, egoistische Handlungen zu unterdrücken und selbstlos zu agieren.

Wir erwarteten für Experiment 4, dass auch das Merkmal Geschlecht des Spielpartners bestimmte Erwartungen auslöst. Von Frauen wird eher kooperatives Verhalten erwartet, so dass mutmaßlich erwartungsinkongruente Defektion von Frauen mehr bestraft wird als Defektion durch Männer. Mit Hilfe des multinomialen Modells wird weiterführend analysiert, ob es sich dabei um einen direkten oder einen indirekten Effekt handelt.

Den Probanden wurden 20 Bilder weiblicher und 20 Bilder männlicher Gesichter aus der »Center for Vital Longevity«-Datenbank (Minear & Park, 2004) präsentiert. Die Bilder waren randomisiert mit Defektion und Kooperation im Spiel gepaart, so dass jede Kombination aus

Geschlecht und Verhalten zehn Mal auftauchte. Wie in den Experimenten 1 und 2 waren die computergesteuerten Spielpartner so programmiert, dass Spielpartner, die zuvor kooperiert hatten, defektierende Probanden randomisiert bestrafen.

Von den 80 Probanden (davon 61 Frauen) bestrafte 67 mindestens einmal eine Defektion und 36 mindestens einmal eine Kooperation. Für den aggregierten Betrag, der für die Bestrafung der Spielpartner ausgegeben wurde (Tabelle 4), reflektierte ein signifikanter Haupteffekt des Verhaltens der Spielpartner [$F(1,79) = 55.24, p < .001, \eta_p^2 = .41$], dass Defektion mehr bestraft wurde als Kooperation. Der Haupteffekt für den Faktor Geschlecht des Spielpartners war nicht signifikant [$F(1,79) = 0.60, p = .439, \eta_p^2 < .01$]. Wie erwartet fand sich eine signifikante Interaktion zwischen dem Verhalten und dem Geschlecht der Spielpartner [$F(1,79) = 7.00, p = .010, \eta_p^2 = .08$]. Die Probanden gaben mehr Geld für die Bestrafung von defektierenden Frauen als für die Bestrafung von defektierenden Männern und mehr Geld für die Bestrafung von kooperierenden Männern als für die Bestrafung von kooperierenden Frauen aus. Die Probanden investierten also mehr Geld in die Bestrafung von erwartungsinkongruentem Verhalten, also von Verhalten, das nicht der Geschlechterrolle entsprach.

Tabelle 4: Aggregierter Betrag, den die Probanden in Experiment 4 ausgaben, um ihre Spielpartner zu bestrafen (summiert über die Durchgänge und gemittelt über die Probanden) in Cent für die Kombinationen aus Geschlecht des Spielpartners (weiblich vs. männlich) und Verhalten des Spielpartners (Defektion vs. Kooperation).

| Geschlecht des Spielpartners | Verhalten des Spielpartners | | | |
|------------------------------|-----------------------------|---------------|-------------|---------------|
| | Defektion | | Kooperation | |
| | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) |
| Weiblich | 15.60 | (1.68) | 2.60 | (0.65) |
| Männlich | 13.45 | (1.48) | 3.95 | (0.94) |

In der multinomialen Analyse passten die Daten gut zum Modell [$G^2(2) = 0.09, p = .957$]. Die Kooperations- und Bestrafungsparameter sind in Abbildung 5 abgebildet. Die Probanden kooperierten mit höherer Wahrscheinlichkeit mit weiblichen als mit männlichen Spielpartnern [$\Delta G^2(1) = 23.99, p < .001, w = 0.09$].

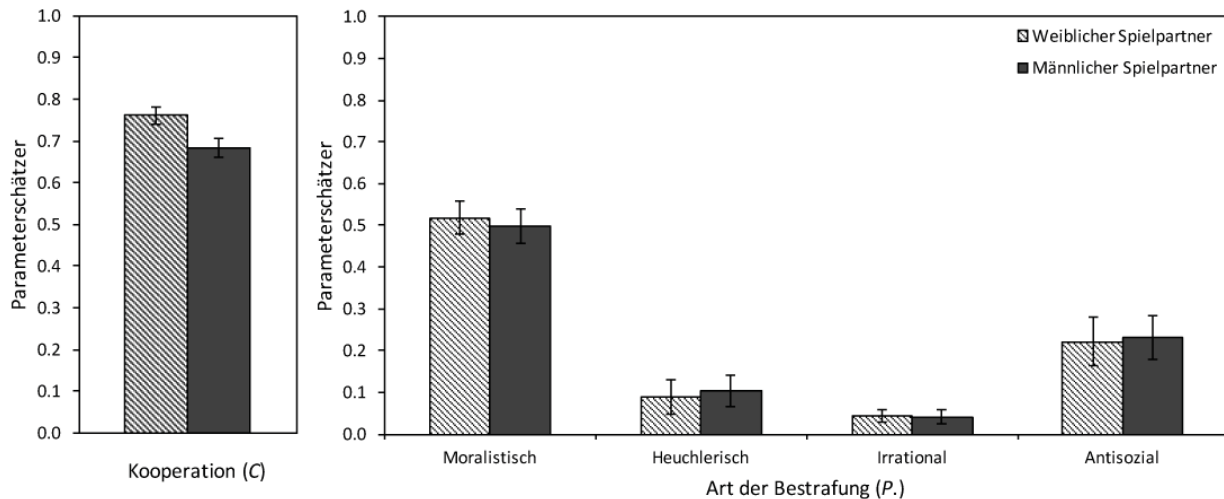


Abbildung 5. Parameterschätzer für Experiment 4. In der linken Grafik sind die Kooperationsparameter (C) abgebildet. Sie repräsentieren die Wahrscheinlichkeiten, dass die Probanden kooperiert haben in Abhängigkeit vom Geschlecht der Spielpartner (weiblich vs. männlich). In der rechten Grafik sind die Bestrafungsparameter (P.) in Abhängigkeit von der Art der Bestrafung (moralistisch, heuchlerisch, irrational und antisozial) und des Geschlechts der Spielpartner (weiblich vs. männlich) abgebildet. Die Fehlerbalken repräsentieren die .95-Konfidenzintervalle.

Moralistische Bestrafung fand sich signifikant wahrscheinlicher als heuchlerische Bestrafung [$\Delta G^2(2) = 257.49, p < .001, w = 0.28$]. Irrationale Bestrafung wurde am seltensten gezeigt und trat damit weniger wahrscheinlich auf als antisoziale Bestrafung [$\Delta G^2(2) = 114.94, p < .001, w = 0.19$].

Weiterhin ergab sich kein Unterschied in der Wahrscheinlichkeit der Bestrafung von weiblichen und männlichen Spielpartnern, weder für moralistische Bestrafung [$\Delta G^2(1) = 0.38, p = .539, w = 0.01$], noch für heuchlerische [$\Delta G^2(1) = 0.20, p = .652, w < 0.01$], irrationale [$\Delta G^2(1) = 0.05, p = .825, w < 0.01$] oder antisoziale Bestrafung [$\Delta G^2(1) = 0.07, p = .790, w < 0.01$]. Auch das Geschlecht des Spielpartners (Tabelle 2 in Mieth, Buchner, & Bell, 2016) wirkte sich nicht direkt auf die Härte der Bestrafung aus [$F(1,62) = 0.19, p = .668, \eta_p^2 < .01$].

Analog zu den vorherigen Experimenten fanden wir einen Effekt des Geschlechts auf den aggregierten Betrag, der für Bestrafung ausgegeben wurde. Hierbei wurde für die Bestrafung von defektierenden Frauen mehr Geld ausgegeben als für Bestrafung von defektierenden Männern. Auf den ersten Blick spricht dies dafür, dass das wohlwollende Stereotyp, Frauen seien freundlicher und weniger kompetitiv als Männer den negativen Effekt mit sich zieht, dass Frauen mehr bestraft werden, wenn sie sich erwartungskongruent verhalten. Der Effekt

beruhte jedoch nicht auf einem direkten Effekt auf die Wahrscheinlichkeit oder Härte der Bestrafung. Vielmehr ergab sich ein indirekter Effekt: Die Probanden kooperierten wahrscheinlicher mit weiblichen als mit männlichen Spielpartnern. In der Kombination mit der Tatsache, dass moralistische Bestrafung wahrscheinlicher gezeigt wird als heuchlerische Bestrafung, bot sich somit mehr Gelegenheit für moralistische Bestrafung.

Experiment 5

In Experiment 4 wurde jedoch der Einfluss, den das Geschlecht der Probanden auf die Interaktion mit den Spielpartnern haben kann, außer Acht gelassen. Dies wurde in den Experimenten 5 und 6 untersucht. Dabei wurden gleich viele männliche und weibliche Probanden erhoben. Da es sich bei den vorherigen Experimenten um überwiegend weibliche Stichproben handelte, konnte der Einfluss des Geschlechts der Probanden in diesen Experimenten nicht sinnvoll untersucht werden. In Experiment 5 untersuchten wir, ob sich die Befunde aus Experiment 4 auch auf männliche Probanden übertragen lassen. Weiterhin sollten in Experiment 5 die Befunde der vorherigen Studie mit einem neuen Stimulusset repliziert werden. Wir erwarteten, die Ergebnisse von Experiment 4 in der Gruppe der weiblichen Probanden zu replizieren: Weibliche Probanden sollten mehr mit weiblichen als mit männlichen Partnern kooperieren. Außerdem sollten sie weibliche und männliche Spielpartner gleich stark moralistisch bestrafen. Im Bezug auf männliche Probanden war offen, wie sie weibliche und männliche Partner behandeln würden.

Betrachtet man Kooperations- und Bestrafungsverhalten für weibliche und männliche Probanden getrennt, sind durchaus Unterschiede zu erwarten. Falls Männer tatsächlich tendenziell eher an der Maximierung ihres eigenen Gewinns interessiert sind und Frauen sich eher für soziale Aspekte einer Aufgabe interessieren (Kahn, Hottes, & Davis, 1971), ist damit zu rechnen, dass einerseits Frauen mehr kooperieren als Männer und dass andererseits Frauen eher bereit sind, Geld für Bestrafung zu investieren. Dieser erste Aspekt konnte in Experiment 5 noch nicht geklärt werden. Die Probanden wurden auch in Experiment 5 für Defektion durch ihre kooperativen Spielpartner moralistisch bestraft, so dass sie aus dem egoistischen Grund, moralistische Bestrafung vermeiden zu wollen, kooperiert haben könnten. Eine reine Kooperationsneigung konnte daher in Experiment 5 noch nicht erfasst werden. Die Auswirkung moralistischer Bestrafung durch die Spielpartner im Bezug auf Geschlechtseffekte wurde in Experiment 6 untersucht. Da Bestrafung in allen vorliegenden Ex-

perimenten mit Kosten für die Probanden verbunden war, konnte jedoch schon in Experiment 5 ein Geschlechtseffekt auf das Bestrafungsverhalten untersucht werden.

Den Probanden wurden erneut jeweils 20 männliche und 20 weibliche Gesichter aus der »Center for Vital Longevity«-Datenbank (Minear & Park, 2004) präsentiert. Die Hälfte der Gesichter war im Spiel jeweils mit Kooperation und die andere Hälfte mit Defektion assoziiert. Wie in den Experimenten 1, 2 und 4 waren die computergesteuerten Spielpartner so programmiert, dass Spielpartner, die zuvor kooperiert hatten, defektierende Probanden randomisiert bestrafte.

Tabelle 5: Aggregierter Betrag, den die Probanden in Experiment 5 ausgaben, um ihre Spielpartner zu bestrafen (summiert über die Durchgänge und gemittelt über die Probanden) in Cent für die Kombinationen aus Geschlecht des Spielpartners (weiblich vs. männlich) und Verhalten des Spielpartners (Defektion vs. Kooperation). Die Ergebnisse werden getrennt für weibliche und männliche Probanden berichtet.

| Geschlecht des Spielpartners | Geschlecht des Probanden | Verhalten des Spielpartners | | | |
|------------------------------|--------------------------|-----------------------------|---------------|-------------|---------------|
| | | Defektion | | Kooperation | |
| | | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) |
| Weiblich | Weiblich | 20.84 | (1.99) | 4.00 | (1.15) |
| Männlich | Weiblich | 19.36 | (1.88) | 4.11 | (1.11) |
| Weiblich | Männlich | 17.23 | (1.70) | 2.75 | (0.81) |
| Männlich | Männlich | 16.76 | (1.74) | 4.23 | (1.22) |

Defektion wurde von 75 der 80 weiblichen Probanden und von 73 der 80 männlichen Probanden mindestens einmal bestraft, während Kooperation von 34 weiblichen und von 29 männlichen Probanden mindestens einmal bestraft wurde. Wie in Experiment 4 fand sich auch in Experiment 5 für den aggregierten Betrag, der für Bestrafung ausgegeben wurde (Tabelle 5), ein signifikanter Haupteffekt für das Verhalten der Spielpartner [$F(1,158) = 204.88, p < .001, \eta_p^2 = .56$]: Es wurde mehr Geld für die Bestrafung von defektierenden als für die Bestrafung von kooperierenden Spielpartnern ausgegeben. Der Haupteffekt des Geschlechts der Spielpartner war nicht signifikant [$F(1,158) = 0.04, p = .847, \eta_p^2 < .01$]. Die Interaktion zwischen Verhalten und Geschlecht der Spielpartner wurde hingegen signifikant [$F(1,158) = 4.83, p = .029, \eta_p^2 = .03$]: Frauen wurden mehr für Defektion bestraft als Männer, während Männer mehr für Kooperation bestraft wurden

als Frauen. Der Haupteffekt der Zwischensubjektvariable Geschlecht des Probanden wurde nicht signifikant, wie auch sämtliche Interaktionen mit diesem Faktor (siehe Mieth, Buchner, et al., 2016, Experiment 2, Results).

In der multinomialen Analyse fand sich eine gute Passung der Daten zum Modell [$G^2(4) = 2.62, p = .624$]. Die Kooperations- und Bestrafungsparameter sind in Abbildung 6 abgebildet. Weibliche Probanden kooperierten signifikant wahrscheinlicher mit weiblichen Spielpartnern als mit männlichen Spielpartnern [$\Delta G^2(1) = 7.03, p = .008, w = 0.03$], während männliche Probanden keinen Unterschied in ihrer Kooperation gegenüber weiblichen und männlichen Spielpartnern machten [$\Delta G^2(1) = 0.61, p = .436, w < 0.01$].

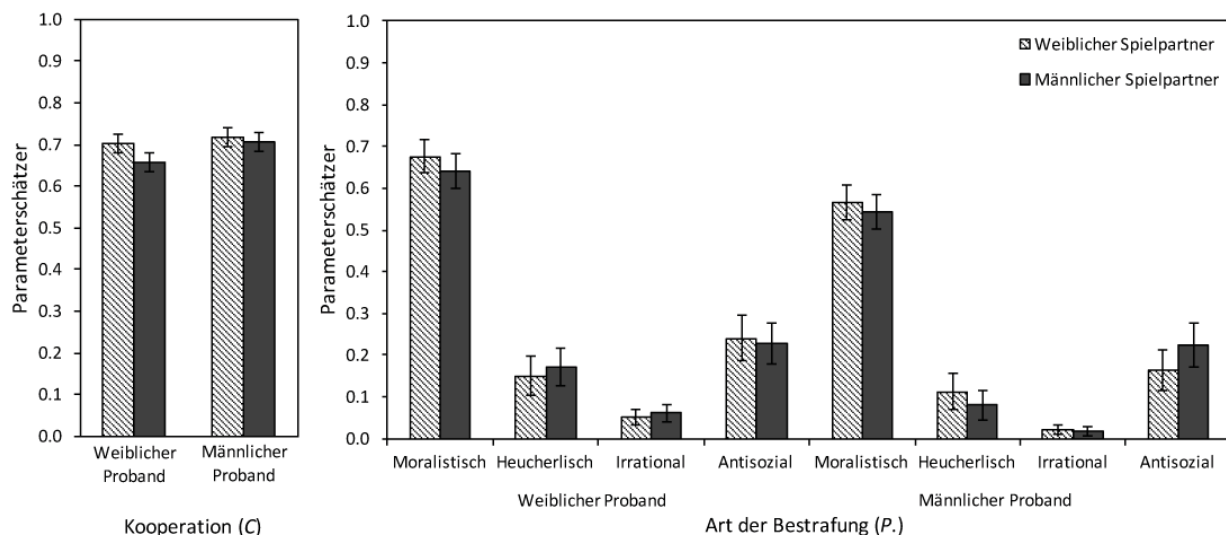


Abbildung 6. Parameterschätzer für Experiment 5. In der linken Grafik sind die Kooperationsparameter (C) abgebildet. Sie repräsentieren die Wahrscheinlichkeiten, dass die Probanden kooperiert haben in Abhängigkeit vom Geschlecht der Spielpartner (weiblich vs. männlich) und dem Geschlecht der Probanden (weiblich vs. männlich). In der rechten Grafik sind die Bestrafungsparameter (P.) in Abhängigkeit von der Art der Bestrafung (moralistisch, heuchlerisch, irrational und antisozial), dem Geschlecht der Spielpartner (weiblich vs. männlich) und dem Geschlecht der Probanden (weiblich vs. männlich) abgebildet. Die Fehlerbalken repräsentieren die .95-Konfidenzintervalle.

In den Bestrafungsparametern zeigte sich, dass moralistische Bestrafung wahrscheinlicher auftrat als heuchlerische Bestrafung [$\Delta G^2(4) = 682.75, p < .001, w = 0.33$]. Irrationale Bestrafung zeigte sich mit geringster Wahrscheinlichkeit und damit seltener als antisoziale Bestrafung [$\Delta G^2(4) = 243.46, p < .001, w = 0.20$].

Schließlich gab es keinen Effekt des Geschlechts der Spielpartner auf die Wahrscheinlichkeit moralistischer Bestrafung oder einer anderen Art der Bestrafung, wenn die Bestrafungen als kondi-

tionale Wahrscheinlichkeiten ausgedrückt wurden. Weibliche Probanden bestraften weibliche und männliche Spielpartner gleich wahrscheinlich moralistisch [$\Delta G^2(1) = 1.41, p = .235, w = 0.01$]. Ebenfalls machten sie keinen Unterschied zwischen weiblichen und männlichen Spielpartner, wenn sie heuchlerisch [$\Delta G^2(1) = 0.45, p = .500, w < 0.01$], irrational [$\Delta G^2(1) = 0.50, p = .477, w < 0.01$] oder antisozial bestraften [$\Delta G^2(1) = 0.14, p = .710, w < 0.01$]. Männliche Probanden zeigten parallele Ergebnisse. Sie machten ebenfalls keinen Unterschied in der moralistischen Bestrafung von weiblichen und männlichen Spielpartnern [$\Delta G^2(1) = 0.61, p = .435, w < 0.01$]. Gleiches galt für heuchlerische Bestrafung [$\Delta G^2(1) = 1.43, p = .232, w = 0.01$], irrationale Bestrafung [$\Delta G^2(1) = 0.17, p = .679, w < 0.01$] und antisoziale Bestrafung [$\Delta G^2(1) = 2.77, p = .096, w = 0.02$].

Weibliche Probanden bestraften wahrscheinlicher als männliche Probanden moralistisch [$\Delta G^2(2) = 25.69, p < .001, w = 0.06$], heuchlerisch [$\Delta G^2(2) = 11.33, p = .003, w = 0.04$] und irrational [$\Delta G^2(2) = 21.83, p < .001, w = 0.06$]. Lediglich für die Wahrscheinlichkeiten antisozialer Bestrafung ergab sich kein Unterschied zwischen weiblichen und männlichen Probanden [$\Delta G^2(2) = 4.50, p = .105, w = 0.03$].

Für die Härte der Bestrafung (Tabelle 2 in Mieth, Buchner, et al., 2016) ergab sich kein signifikanter Haupteffekt des Zwischensubjektfaktors Geschlecht der Probanden [$F(1,133) = 0.01, p = .912, \eta_p^2 < .01$]. Auch der Haupteffekt des Geschlechts der Spielpartner [$F(1,133) = 1.73, p = .191, \eta_p^2 = .01$] sowie die Interaktion zwischen Geschlecht des Probanden und Geschlecht des Spielpartners wurde nicht signifikant [$F(1,133) = 1.08, p = .301, \eta_p^2 < .01$]. Demnach bestraften Probanden unabhängig von ihrem Geschlecht gleich hart. Das Geschlecht der Spielpartner wirkte sich ebenfalls nicht auf die Härte der Bestrafung aus.

Somit ergab sich also für weibliche Probanden ein paralleles Muster zu den Ergebnissen aus Experiment 4. Sie kooperierten eher mit weiblichen Spielpartnern als mit männlichen, aber bestraften weibliche und männliche Spielpartner gleich wahrscheinlich moralistisch. In der ebenso großen Stichprobe mit männlichen Probanden zeigte sich ein davon verschiedenes Ergebnismuster. Männliche Probanden kooperierten nicht unterschiedlich, sondern gleich wahrscheinlich mit weiblichen und männlichen Spielpartnern. Genau wie weibliche Probanden machten auch männliche Probanden keinen Unterschied in der moralistischen oder irgendeiner anderen Art der Bestrafung

weiblicher und männlicher Spielpartner. Somit bestätigte sich, dass die Effekte von Geschlecht auf Bestrafung indirekt sind.

Interessanterweise bestrafte Frauen vermehrt moralistisch im Vergleich zu Männern. Da moralistische Bestrafung mit Kosten für die Probanden verbunden war, sich aber keine direkten Vorteile für die Probanden ergaben, passt dies gut zu der Annahme, dass Männer eher an der eigenen Gewinnmaximierung interessiert sind, während Frauen eher durch interpersonelle Aspekte von sozialen Dilemmaspielen beeinflusst werden (Andreoni & Vesterlund, 2001; Eckel & Grossman, 1998, 2008; Kahn et al., 1971). Frauen zeigten auch wahrscheinlicher heuchlerische und irrationale Bestrafung. Nur für die Wahrscheinlichkeit antisozialer Bestrafung fand sich kein Unterschied zwischen den Geschlechtern der Probanden.

Experiment 6

Was schon in Bezug auf die Sympathie des Gesichts der Spielpartner galt, könnte auch in den Untersuchungen zum Einfluss der Geschlechter der Probanden und Spielpartner auf das Kooperations- und Bestrafungsverhalten bestätigt werden: Das Bestrafungsverhalten der Spielpartner beeinflusst das Verhalten der Probanden. In den Experimenten 4 und 5 bestrafte äquivalent zu den Experimenten 1 und 2 kooperative Spielpartner defektierende Probanden, so wie es üblicherweise in der Interaktion von echten Menschen geschieht (Carpenter, 2007; Falk et al., 2005). Diese durch die Spielpartner ausgeübte moralistische Bestrafung könnte das Kooperations- wie auch das Bestrafungsverhalten der Probanden beeinflusst haben, da es zum einen verstärkend auf die Kooperationsnorm gewirkt haben könnte und zum anderen das Bestrafungsverhalten implizit von den Spielpartnern gelernt worden sein könnte. Darüber hinaus untersuchten wir, ob sich das Bestrafungsverhalten der Spielpartner anders auf männliche als auf weibliche Probanden auswirkt. In Experiment 6 stand die Bestrafungsoption deshalb wie Experiment 3 ausschließlich den Probanden zur Verfügung – die Spielpartner bestrafte die Probanden nicht.

Moralistische Bestrafung soll kooperatives Verhalten fördern, da die Bestrafung von unkooperativem Verhalten antizipiert wird. Demnach war für Experiment 6 wie schon für Experiment 3 zu erwarten, dass die Kooperation der Probanden in dem Maß zurückgehen würde, in dem sie Bestrafung durch ihre Spielpartner erwartet und nur aus Angst davor kooperiert hatten. Unklar ist,

warum weibliche Probanden in Experiment 5 mehr mit weiblichen Spielpartnern kooperierten als mit männlichen. Dies könnte sich ergeben, wenn Frauen eher motiviert sind, mit weiblichen Spielpartnern zu kooperieren, weil sie von diesen eher reziproke Kooperation erwarteten oder weil die kooperative Handlung als solche belohnend wirkte. Allerdings könnten Frauen ebenso die moralistische Bestrafung weiblicher Spielpartner mehr fürchteten als die moralistische Bestrafung männlicher Spielpartner und daher eher mit weiblichen als mit männlichen Spielpartnern kooperieren.

Außerdem interessierte uns, warum weibliche Probanden im Vergleich zu männlichen Probanden mehr bestrafte. Durch die moralistische Bestrafung der Spielpartner könnte eine Bestrafungsnorm bei den Probanden entstanden sein, die dazu geführt haben könnte, dass sie wahrscheinlicher ebenfalls moralistisch bestrafte, obwohl sie dieses Verhalten ohne Vorbilder vielleicht nicht gezeigt hätten. In diesem Falle scheint es plausibel, dass weibliche Probanden nicht aufgrund intrinsischer Motivation vermehrt moralistisch bestrafte, sondern, weil sie sich stärker durch das Bestrafungsverhalten ihrer Partner beeinflussen ließen als männliche Probanden und somit die Norm, moralistisch zu bestrafen, adaptierten.

Experiment 6 war identisch mit Experiment 5 mit der einzigen Ausnahme, dass nur den Probanden die Bestrafungsoption zur Verfügung stand. Von den 81 weiblichen Probanden nutzten 63 diese Option und bestrafte mindestens einmal eine Defektion. Von den 80 männlichen Probanden bestrafte 50 mindestens einmal eine Defektion. Außerdem bestrafte 24 weibliche und 17 männliche Probanden mindestens einmal eine Kooperation. Für den aggregierten Geldbetrag, der für die Bestrafung der Spielpartner ausgegeben wurde (Tabelle 6), ergab sich ein signifikanter Haupteffekt des Verhaltens der Spielpartner [$F(1,159) = 73.90, p < .001, \eta_p^2 = .32$]: Defektierende Spielpartner wurden mehr bestrafte als kooperierende. Weder der Haupteffekt des Geschlechts der Spielpartner [$F(1,159) = 3.08, p = .081, \eta_p^2 = .02$], noch die Interaktion zwischen Geschlecht und Verhalten des Spielpartners war signifikant [$F(1,159) = 0.09, p = .761, \eta_p^2 < .01$]. Gleiches galt auch für den Zwischensubjekteffekt Geschlecht des Probanden [$F(1,159) = 1.19, p = .276, \eta_p^2 < .01$] sowie für sämtliche Interaktionen mit diesem Faktor (siehe Mieth, Buchner, et al., 2016, Experiment 3, Results).

Tabelle 6: Aggregierter Betrag, den die Probanden in Experiment 6 ausgaben, um ihre Spielpartner zu bestrafen (summiert über die Durchgänge und gemittelt über die Probanden) in Cent für die Kombinationen aus Geschlecht des Spielpartners (weiblich vs. männlich) und Verhalten des Spielpartners (Defektion vs. Kooperation). Die Ergebnisse werden getrennt für weibliche und männliche Probanden berichtet.

| Geschlecht des Spielpartners | Geschlecht des Probanden | Verhalten des Spielpartners | | | |
|------------------------------|--------------------------|-----------------------------|---------------|-------------|---------------|
| | | Defektion | | Kooperation | |
| | | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) |
| Weiblich | Weiblich | 8.26 | (0.96) | 1.95 | (0.59) |
| Männlich | Weiblich | 8.15 | (1.05) | 2.42 | (0.69) |
| Weiblich | Männlich | 6.73 | (1.28) | 0.63 | (0.22) |
| Männlich | Männlich | 8.05 | (1.48) | 1.01 | (0.37) |

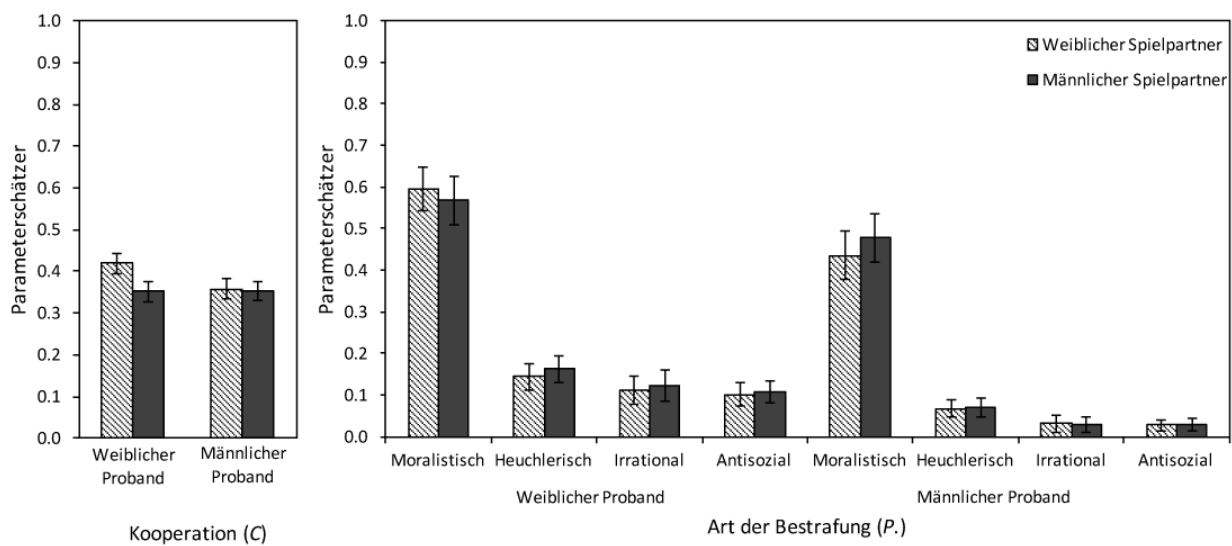


Abbildung 7. Parameterschätzer für Experiment 6. In der linken Grafik sind die Kooperationsparameter (*C*) abgebildet. Sie repräsentieren die Wahrscheinlichkeiten, dass die Probanden kooperiert haben in Abhängigkeit vom Geschlecht der Spielpartner (weiblich vs. männlich) und dem Geschlecht der Probanden (weiblich vs. männlich). In der rechten Grafik sind die Bestrafungsparameter (*P.*) in Abhängigkeit von der Art der Bestrafung (moralistisch, heuchlerisch, irrational und antisozial), dem Geschlecht der Spielpartner (weiblich vs. männlich) und dem Geschlecht der Probanden (weiblich vs. männlich) abgebildet. Die Fehlerbalken repräsentieren die .95-Konfidenzintervalle.

In der multinomialen Analyse passten die Daten gut zum Modell [$G^2(4) = 1.71, p = .788$]. Die Kooperations- und Bestrafungsparameter sind in Abbildung 7 abgebildet. Wie in Experiment 5 kooperierten weibliche Probanden wahrscheinlicher mit weiblichen als mit männlichen Spielpartnern [$\Delta G^2(1) = 15.50, p < .001, w = 0.05$], während männliche Probanden in ihrem Koopera-

tionsverhalten keinen Unterschied zwischen weiblichen und männlichen Spielpartnern machten [$\Delta G^2(1) = 0.07, p = .796, w < 0.01$].

Moralistisches Bestrafungsverhalten war wahrscheinlicher als heuchlerisches [$\Delta G^2(4) = 653.32, p < .001, w = 0.32$]. Irrationale Bestrafung hingegen trat anders als in den Experimenten 4 und 5, aber parallel zu Experiment 3 genauso wahrscheinlich auf wie antisoziale Bestrafung [$\Delta G^2(4) = 0.82, p = .936, w = 0.01$]. Dies passt wiederum zu Experiment 3 und legt nahe, dass antisoziale Bestrafung einen Präventivschlag darstellt, wenn von einem kooperativen Spielpartner moralistische Bestrafung antizipiert wird (Nikiforakis, 2008).

Wie in den Experimenten 4 und 5 fanden sich keine direkten Effekte des Geschlechts des Spielpartners auf die Wahrscheinlichkeit für moralistische Bestrafung oder eine andere Art der Bestrafung, wenn diese als konditionale Wahrscheinlichkeiten ausgedrückt wurden. Frauen bestrafte weder weibliche noch männliche Spielpartnern wahrscheinlicher moralistisch [$\Delta G^2(1) = 0.45, p = .503, w < 0.01$]. Gleiches galt für heuchlerische Bestrafung [$\Delta G^2(1) = 0.66, p = .415, w = 0.01$], irrationale Bestrafung [$\Delta G^2(1) = 0.18, p = .670, w < 0.01$] und antisoziale Bestrafung [$\Delta G^2(1) = 0.05, p = .820, w < 0.01$]. Männer zeigten ein analoges Befundmuster und bestrafte ebenfalls weibliche und männliche Spielpartner gleich wahrscheinlich für moralistische Bestrafung [$\Delta G^2(1) = 1.07, p = .301, w = 0.01$], heuchlerische Bestrafung [$\Delta G^2(1) = 0.04, p = .848, w < 0.01$], irrationale Bestrafung [$\Delta G^2(1) = 0.02, p = .893, w < 0.01$] und antisoziale Bestrafung [$\Delta G^2(1) = 0.01, p = .913, w < 0.01$].

Parallel zum vorherigen Experiment bestrafte weibliche Probanden wahrscheinlicher als männliche Probanden moralistisch [$\Delta G^2(2) = 20.87, p < .001, w = 0.06$], heuchlerisch [$\Delta G^2(2) = 37.89, p < .001, w = 0.08$], irrational [$\Delta G^2(2) = 34.54, p < .001, w = 0.07$] und auch antisoziale [$\Delta G^2(2) = 51.24, p < .001, w = 0.09$].

Für die Härte der Bestrafung (Tabelle 2 in Mieth, Buchner, et al., 2016) ergab sich konsistent zu den vorherigen Untersuchungen kein signifikanter Haupteffekt für den Faktor Geschlecht des Spielpartners [$F(1,91) = 0.78, p = .379, \eta_p^2 < .01$] und auch der Zwischensubjekteffekt Geschlecht der Probanden verfehlte das Signifikanzniveau [$F(1,91) = 3.59, p = .061, \eta_p^2 = .04$]. Die Interak-

tion zwischen den Faktoren Geschlecht der Probanden und Geschlecht der Spielpartner war ebenfalls nicht signifikant [$F(1,91) = 1.15, p = .286, \eta_p^2 = .01$].

Damit fanden wir in Experiment 6 parallele Ergebnisse zu Experiment 5. Weibliche Probanden kooperierten mehr mit weiblichen als mit männlichen Spielpartnern, während männliche Probanden keinen Unterschied in der Kooperation mit weiblichen und männlichen Spielpartnern machten. Da die Spielpartner in Experiment 6 keine Bestrafung ausführten, scheint die Idee widerlegt, dass weibliche Probanden hauptsächlich aus Angst vor der moralistischen Bestrafung ihrer weiblichen Spielpartner kooperierten. Weiterführend zeigten die Probanden abermals moralistische Bestrafung mit höherer Wahrscheinlichkeit als andere Arten der Bestrafung. Dementsprechend liegt die Schlussfolgerung nahe, dass auch das Bestrafungsverhalten in den Experimenten 4 und 5 nicht nur dadurch zustande kam, dass die Probanden das Bestrafungsverhalten ihrer Spielpartner adaptierten. Auch wenn die Probanden sich nicht an ihren Spielpartnern und deren Bestrafungsverhalten orientieren konnten, bestrafte sie in großem Maße moralistisch. Darüber hinaus hatte das Geschlecht der Spielpartner wie in den vorherigen Experimenten keinen direkten Einfluss auf die Wahrscheinlichkeit von Bestrafung: Die Probanden bestrafte weibliche und männliche Spielpartner mit derselben Wahrscheinlichkeit. Schließlich konnten wir den Befund replizieren, dass weibliche Probanden wahrscheinlicher moralistisch bestrafen als männliche Probanden.

Nichtsdestotrotz ergaben sich auch Unterschiede zu Experiment 5. Wie in Experiment 3 im Vergleich zu Experiment 2 reduzierte sich die Kooperationsbereitschaft in Experiment 6 im Vergleich zu Experiment 5. Moralistische Bestrafung fördert also die Kooperation. Daher war die Kooperation der Probanden in den Experimenten 4 und 5 sicherlich in Teilen dadurch motiviert, dass sie Angst vor der moralistischen Bestrafung ihrer Spielpartner hatten. Da die Kooperationsbereitschaft in Experiment 6 jedoch nicht in Gänze verschwand, ist davon auszugehen, dass Kooperation nicht ausschließlich durch Angst vor Bestrafung motiviert war. In Experiment 6 konnten die Probanden ihren Gewinn maximieren, indem sie die Kooperation verweigerten, ohne dass dies bestraft wurde. Damit gab es keinen monetären Anreiz für Kooperation. Um den eigenen Gewinn zu maximieren, sollten sich die Probanden rationalerweise dafür entscheiden zu defektieren. Dennoch ergaben sich relativ hohe Wahrscheinlichkeiten für Kooperation, obwohl keine Bestrafung durch die Spielpartner drohte.

Wenn die Ergebnisse zweier verschiedener Experimente auch schwer direkt miteinander zu vergleichen sind, ist festzustellen, dass der Rückgang der Kooperation von Experiment 5 zu Experiment 6 für männliche Probanden ausgeprägter zu sein scheint, als für weibliche Probanden. Dieser Unterschied zwischen den Experimenten könnte darauf zurückzuführen sein, dass kooperatives Verhalten männlicher Probanden stärker auf monetären Anreizen basierte und damit sensitiver für moralistische Bestrafung war als kooperatives Verhalten weiblicher Probanden (Andreoni & Vesterlund, 2001; Eckel & Grossman, 1998, 2008). Weibliche Probanden könnten hingegen eher motiviert gewesen sein zu kooperieren, weil eine kooperative Handlung als solche belohnend wirkt.

Obwohl sich wiederum hohe Wahrscheinlichkeiten für moralistische Bestrafung fanden, sind Unterschiede zu den vorherigen Experimenten erkennbar. Die Wahrscheinlichkeit antisozialer Bestrafung war in den Experimenten 4 und 5 deutlich höher als in Experiment 6 und unterschied sich hier nicht mehr signifikant von der Wahrscheinlichkeit für irrationale Bestrafung. Dies passt wiederum zu Experiment 3 und der Erklärung, dass antisoziale Bestrafung durch die Vergeltung antizipierter moralistischer Bestrafung durch die Spielpartner motiviert sein könnte (Nikiforakis, 2008).

Allgemeine Diskussion

In der vorliegenden Arbeit wurde untersucht, wie sich die Erwartungen an soziale Interaktionspartner auf Kooperations- und Bestrafungsverhalten in einem Gefangenendilemma mit Bestrafungsoption auswirkten. Heuristiken und Stereotype helfen unsere Umwelt zu strukturieren und liefern Handlungsempfehlungen in unsicheren Situationen. Sie können durch verschiedenste Merkmale hervorgerufen werden. Gesichtsmerkmale können beispielsweise helfen eine Kooperations- oder Defektionsentscheidung zu treffen. Positive Stereotype über ein Gegenüber gehen mit entsprechenden Erwartungen einher. Von sympathischen, lächelnden und weiblichen Personen wird eher positives, und damit kooperatives Verhalten erwartet. Verhält sich eine Person nicht erwartungskongruent, resultiert dies in Enttäuschung, die folglich zu verstärkter Bestrafung der betreffenden Person führen kann. Um diese These zu überprüfen, wurden in der vorliegenden Arbeit in sechs Experimenten für die Spielpartner in einem Gefangenendilemma die Gesichtsmerkmale Sympathie des Gesichts, Gesichtsausdruck und das Geschlecht manipuliert. Weiterhin interessierte der Einfluss des Geschlechts der Probanden. Wir analysierten, ob ein Effekt erhöhter Bestrafung

erwartungsinkongruenter Defektion auf einen direkten Effekt erhöht wahrscheinlicher oder härterer Bestrafung oder auf einen indirekten Effekt erhöht wahrscheinlicher Kooperation zurückzuführen ist. In den vorliegenden Experimenten spielten die Probanden mit fiktiven Spielpartnern ein Gefangenendilemma, auf das eine direkte Bestrafungsoption folgte. Eine multinomiale Analyse erlaubte die differenzierende Betrachtung von Kooperation und Bestrafung, indem Bestrafung als konditionale Wahrscheinlichkeit ausgedrückt wurde. So konnten die zugrundeliegenden Prozesse des Bestrafungsverhaltens untersucht werden.

Wir überprüften in den vorliegenden Experimenten zunächst, ob sympathische im Vergleich zu unsympathischen Personen, lächelnde im Vergleich zu neutral schauenden Personen sowie Frauen im Vergleich zu Männern als Folge von unkooperativem Verhalten stärker bestraft werden. In einer ersten aggregierten Analyse des Betrags, der für die Bestrafung von Spielpartnern ausgegeben wird, wurden tatsächlich mehr Ressourcen aufgewendet, um sympathische, lächelnde oder weibliche Spielpartner zu bestrafen im Vergleich zu unsympathischen, neutral schauenden oder männlichen Spielpartnern (Experimente 1, 2 und 4). Somit bestätigten wir zunächst für die verschiedenen Gesichtsmerkmale, dass erwartungsinkongruente Defektion zu erhöhter Bestrafung führt. Die Erwartungen an die Spielpartner beeinflussten jedoch neben dem Bestrafungsverhalten auch die Kooperationsbereitschaft diesen Spielpartnern gegenüber. Um die Kooperationsbereitschaft und das Bestrafungsverhalten der Probanden getrennt voneinander zu untersuchen, verwendeten wir ein multinomiales Modell. Während auf den ersten Blick sympathische, lächelnde oder weibliche Spielpartner mehr bestraft wurden als unsympathische, neutral schauende oder männliche Spielpartner, zeigte sich weiterhin, dass es sich dabei nicht um einen direkten Effekt auf die Wahrscheinlichkeit oder Härte der Bestrafung handelte.

Vielmehr fand sich ein indirekter Effekt: Positive Erwartungen gegenüber sympathischen, lächelnden oder weiblichen Spielpartnern erhöhten die Bereitschaft, mit diesen zu kooperieren. Dies führte zu mehr Gelegenheit lächelnde, sympathische oder weibliche Spielpartner moralistisch zu bestrafen, da diese häufigste Form der Bestrafung nur auftreten kann, wenn der Proband sich vorher selbst kooperativ verhalten hat. Sympathische, lächelnde oder weibliche Spielpartner erfuhr somit mehr moralistische Bestrafung als unsympathische, neutral schauende oder männliche Spielpartner. Demnach haben sympathische, lächelnde oder weibliche Personen keinen

direkten Nachteil aus der Tatsache, dass positive Stereotype mit diesen Merkmalen verbunden und folglich kooperatives Verhalten erwartet wird – sie werden nicht direkt mehr bestraft als unsympathische, neutral schauende und männliche Personen. Die vorliegenden Experimente stützen im Sinne starker Reziprozität (Gintis, 2000; Gintis et al., 2003) die Funktion von Bestrafung, eine Kooperationsnorm allgemein – allen Spielpartnern gegenüber gleichermaßen – durchzusetzen, weil Bestrafung gleichermaßen allen Spielpartnern gegenüber eingesetzt wird.

Interessanterweise zeigte sich in Experiment 2 für die Härte der Bestrafung lächelnder Spielpartner sogar, dass lächelnde im Vergleich zu neutral schauenden Spielpartnern weniger hart moralistisch bestraft wurden. Dies ist wiederum konsistent mit Befunden von Mussel und Kollegen (Mussel et al., 2013; Mussel et al., 2014), in denen lächelnde Personen weniger bestraft wurden. Weiterhin zeigten sie, dass eine negative neuronale Komponente, die mit dem Erhalten negativen Feedbacks assoziiert ist, weniger stark ausfällt bei unfairen Angeboten von lächelnden im Vergleich zu nicht lächelnden Spielpartnern. Dies ist in Zusammenhang damit zu sehen, dass Lächeln ganz allgemein positive Gefühle im Gegenüber erzeugen kann. Außerdem wirkt Lächeln in sozialen Interaktionen einerseits annähernd und hilft andererseits Aggression abzubauen (Goldenthal et al., 1981; Mackey, 1976).

Moralistische Bestrafung wurde als implizite Norm durch das Bestrafungsverhalten der Spielpartner in den Experimenten 1, 2, 4 und 5 vorgegeben. In den Experimenten 3 und 6 bestraften die Spielpartner die Probanden nicht. Da dies sowohl das Kooperations- als auch das Bestrafungsverhalten der Probanden beeinflusst haben könnte, zeigte sich in den Experimenten 3 und 6, dass die Wahrscheinlichkeit für Kooperation seitens der Probanden zurückging, wenn auf ihre Defektion keine moralistische Bestrafung folgte. Außerdem nahm die Wahrscheinlichkeit der Bestrafung ab. Das Befundmuster, dass moralistische Bestrafung am wahrscheinlichsten stattfand, blieb jedoch auch ohne Bestrafung durch die Spielpartner erhalten. Die Probanden kooperierten nicht nur, weil sie durch eine Angst vor der Bestrafung durch die Spielpartner dazu motiviert waren. Weiterhin zeigten die Probanden Bestrafungsverhalten nicht ausschließlich, weil ihre Spielpartner dieses durch eine implizite Bestrafungsnorm vorgaben.

Der Einfluss des Geschlechts der Probanden wurde in den Experimenten 5 und 6 untersucht. In Experiment 6 zeigte sich verglichen mit Experiment 5 ein stärkerer Rückgang von Kooperations-

und Bestrafungsverhalten für Männer als für Frauen. In den Experimenten 5 und 6 zeigte sich weiterhin, dass Frauen im Gegensatz zu Männern einen Unterschied in der Kooperation mit weiblichen und männlichen Spielpartnern machten. Ebenfalls zeigte sich in Experiment 6 im Vergleich zu Experiment 5 ein stärkerer Rückgang von Kooperations- und Bestrafungsverhalten für Männer als für Frauen. Weiterhin nutzten Frauen in den Experimenten 5 und 6 wahrscheinlicher die moralistische Bestrafungsoption (wie auch alle anderen Arten der Bestrafung) als Männer. Dies kann dafür sprechen, dass Männer im Rahmen dieser sozialen Dilemmasituation rationaler und analytischer agiert und weniger auf die sozialen Aspekte fokussiert haben, während Frauen eher an der Aufrechterhaltung von kooperativem Verhalten interessiert waren und dafür auch persönlich investierten (Andreoni & Vesterlund, 2001; Eckel & Grossman, 1998, 2008). Es ist möglich, aus den Experimenten die Schlussfolgerung zu ziehen, dass auch in den Experimenten 1–3 dadurch, dass die Probandenstichproben mehrheitlich Frauen beinhaltete, vor allem Effekte weiblichen Kooperations- und Bestrafungsverhaltens untersucht werden konnten. Demnach bleibt offen, ob der Effekt, dass männliche Probanden in ihrem Kooperations- und Bestrafungsverhalten weniger zwischen weiblichen und männlichen Spielpartnern differenzieren, gegebenenfalls auf weitere Gesichtsm Merkmale wie sympathisch und unsympathisch aussehende oder lächelnde und neutral schauende Spielpartner generalisiert werden kann. Es erscheint möglich, dass Männer eher auf rationale Aspekte der Spielsituation fokussieren und dadurch ihre Kooperationsentscheidungen gegebenenfalls seltener als Frauen auf Aussehensmerkmale beziehen, die Kooperation oder Defektion ihres Gegenübers nahe legen.

In den vorliegenden Experimenten interagierten die Probanden nicht mit realen Personen, sondern mit programmierten Spielpartnern. Dies sorgte für ein hohes Maß an experimenteller Kontrolle der Erfahrungen der Probanden. Dennoch wurde um echtes Geld gespielt, das den Probanden am Ende ausbezahlt wurde. A priori war jedoch nicht klar, ob die Probanden überhaupt bestrafen würden und ob sie in den Experimenten 3 und 6 kooperieren würden. Es zeigte sich jedoch, dass die Probanden durchaus kooperativ agierten und darüber hinaus bereit waren, Geld für die Bestrafung ihrer Spielpartner einzusetzen, obwohl sie nicht davon profitieren konnten. Es fanden sich ähnlich hohe Bestrafungsraten wie in Untersuchungen, in denen mit echten Menschen interagiert wurde (Falk et al., 2005). Und auch in den Experimenten 3 und 6 kooperierten die Probanden zu einem beachtlichen Anteil, obwohl Defektion die eher Gewinn bringende Strategie darstellte und

keine Bestrafung zu befürchten war. Aus ökonomischer Perspektive ist solches Verhalten nicht nachvollziehbar, jedoch ist es absolut vergleichbar mit Verhalten, welches gegenüber realen Interaktionspartnern gezeigt wird.

Ausblickend lässt sich feststellen, dass die hier untersuchten Heuristiken wie die Beurteilung der Sympathie des Gesichts, das Heranziehen eines Lächelns als Aussicht auf kooperatives Verhalten sowie die geschlechterstereotype positive Beurteilung von Frauen allesamt eher sozial akzeptierte Stereotype darstellen. Es erscheint plausibel, dass Probanden ihr Kooperations- und Bestrafungsverhalten bei der experimentellen Vorgabe von sozial nicht akzeptierten Stereotypen oder Vorurteilen anpassen würden. Es könnte daher interessant sein zu untersuchen, wie Probanden sich verhalten, wenn sie sich Personen gegenüber sehen, für die weniger sozial akzeptierte Stereotype gelten.

Zusammenfassend erschien die Annahme, dass erwartungsinkongruente im Vergleich zu erwartungskongruenter Defektion besonders wahrscheinlich oder besonders hart bestraft wird, zunächst plausibel, konnte in den vorliegenden sechs Experimenten jedoch nicht bestätigt werden. Es fanden sich vielmehr indirekte Effekte, dass – vor allem weibliche – Probanden mit sympathischen, lächelnden oder weiblichen Spielpartnern stereotypkonform wahrscheinlicher kooperierten als mit unsympathischen, neutral schauenden oder männlichen Spielpartnern, so dass sich mehr Gelegenheit für moralistische Bestrafung bot, wenn ihre Kooperation nicht erwidert wurde.

Literatur

- Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.
- Andreoni, J., & Vesterlund, L. (2001). Which is the Fair Sex? Gender Differences in Altruism. *The Quarterly Journal of Economics*, *116*, 293–312.
- Ashmore, R., & Del Boca, F. (1979). Sex stereotypes and implicit personality theory: Toward a cognitive—social psychological conceptualization. *Sex Roles*, *5*, 219–248.
- Balliet, D., Li, N., Macfarlan, S., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin*, *137*, 881–909.
- Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 197–215.
- Bell, R., Buchner, A., Kroneisen, M., & Giang, T. (2012). On the flexibility of social source memory: A test of the emotional incongruity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1512–1529.
- Bell, R., Buchner, A., & Musch, J. (2010). Enhanced old-new recognition and source memory for faces of cooperators and defectors in a social-dilemma game. *Cognition*, *117*, 261–275.
- Bell, R., Giang, T., Mund, I., & Buchner, A. (2013). Memory for reputational trait information: Is social–emotional information processing less flexible in old age? *Psychology and Aging*, *28*, 984–995.
- Bell, R., Sasse, J., Möller, M., Czernochowski, D., Mayr, S., & Buchner, A. (2016). Event-related potentials in response to cheating and cooperation in a social dilemma game. *Psychophysiology*, *53*, 216–228.
- Bernstein, F. (1925). Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen [Summarizing consideration on the inheritable blood structure of mankind]. *Zeitschrift für Abstammungs- und Vererbungslehre*, *37*, 237–270.
- Bonnefon, J.-F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General*, *142*, 143–150.
- Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie/Journal of Psychology*, *215*, 52–60.

- Brown, W. M., & Moore, C. (2002). Smile asymmetries and reputation as reliable indicators of likelihood to cooperate: An evolutionary analysis. In S. P. Shohov (Ed.), *Advances in Psychology Research* (pp. 59–78). Huntington, New York.
- Carpenter, J. (2007). The demand for punishment. *Journal of Economic Behavior & Organization*, 62, 522–542.
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24, 285–290.
- Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality & Social Psychology Bulletin*, 15, 543–558.
- Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal*, 108, 726–735.
- Eckel, C. C., & Grossman, P. J. (2008). Differences in the economic decisions of men and women: Experimental evidence. In C. Plott, & V. Smith (Eds.), *Handbook of Experimental Economics Results* (pp. 509–519). New York: Elsevier.
- Eckel, C. C., & Wilson, R. K. (2003). The human face of game theory: Trust and reciprocity in sequential games. In E. Ostrom, & J. M. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 245–274). New York: Sage.
- Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala. *Journal of Cognitive Neuroscience*, 19, 1508–1519.
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. *Zeitschrift für Psychologie / Journal of Psychology*, 217, 108–124.
- Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, 73, 2017–2030.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–140.
- Fetchenhauer, D., Groothuis, T., & Pradel, J. (2010). Not only states but traits — Humans can identify permanent altruistic dispositions in 20 s. *Evolution and Human Behavior*, 31, 80–86.
- Giang, T., Bell, R., & Buchner, A. (2012). Does facial resemblance enhance cooperation? *PLoS ONE*, 7, e47809.

- Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206, 169–179.
- Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24, 153–172.
- Goldenthal, P., Johnston, R., & Kraut, R. (1981). Smiling, appeasement, and the silent bared-teeth display. *Ethology and Sociobiology*, 2, 127–133.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2006). Costly punishment across human societies. *Science*, 312, 1767–1770.
- Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, 319, 1362–1367.
- Kahn, A., Hottes, J., & Davis, W. L. (1971). Cooperation and optimal responding in the Prisoner's Dilemma game: Effects of sex and physical attractiveness. *Journal of Personality and Social Psychology*, 17, 267–279.
- Klauer, K., Stahl, C., & Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 680–703.
- Ma, Q., Hu, Y., Jiang, S., & Meng, L. (2015). The undermining effect of facial attractiveness on brain responses to fairness in the ultimatum game: an ERP study. *Frontiers in Neuroscience*, 9.
- Mackey, W. (1976). Parameters of the smile as a social signal. *The Journal of Genetic Psychology*, 129, 125–130.
- Mieth, L., Bell, R., & Buchner, A. (2016). Facial likability and smiling enhance cooperation, but have no direct effect on moralistic punishment. *Manuscript submitted for publication*.
- Mieth, L., Buchner, A., & Bell, R. (2016). Effects of gender on costly punishment. *Manuscript submitted for publication*.
- Miner, M., & Park, D. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36, 630–633.
- Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, 42, 42–54.
- Mussel, P., Göritz, A. S., & Hewig, J. (2013). The value of a smile: Facial expression affects ultimatum-game responses. *Judgment and Decision Making*, 8, 381–385.

- Mussel, P., Hewig, J., Allen, J., Coles, M., & Miltner, W. (2014). Smiling faces, sometimes they don't tell the truth: Facial expression in the ultimatum game impacts decision making and event-related potentials. *Psychophysiology, 51*, 358–363.
- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics, 92*, 91–112.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science, 314*, 1560–1563.
- Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95*, 318–339.
- Scharlemann, J. P. W., Eckel, C., Kacelnik, A., & Wilson, R. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22*, 617–640.
- Snyder, M., Tanke, E., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*, 656–666.
- Solnick, S., & Schweitzer, M. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Processes, 79*, 199–215.
- Todorov, A. (2008). Evaluating faces on trustworthiness. *Annals of the New York Academy of Sciences, 1124*, 208–224.
- Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition, 77*, 813–833.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*, 592–598.
- Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly, 59*, 189–202.

Einzelarbeiten

Experimente 1–3:

Mieth, L., Bell, R., & Buchner, A. (2016). Facial likability and smiling enhance cooperation, but have no direct effect on moralistic punishment. *Manuscript submitted for publication.*

Experiment 4–6:

Mieth, L., Buchner, A., & Bell, R. (2016). Effects of gender on costly punishment. *Manuscript submitted for publication.*

Facial likability and smiling enhance cooperation, but have no direct effect on moralistic punishment

Laura Mieth

Raoul Bell

Axel Buchner

Heinrich-Heine-Universität Düsseldorf

Running Head: Facial appearance and punishment

Corresponding Author:

Laura Mieth

Institute of Experimental Psychology

Heinrich Heine University Düsseldorf

40225 Düsseldorf, Germany

E-Mail: laura.mieth@hhu.de

Abstract

The present study serves to test how positive and negative appearance-based expectations affect cooperation and punishment. Participants played a prisoner's dilemma game with partners who either cooperated or defected. Then they were given a costly punishment option: They could spend money to decrease the payoffs of their partners. Aggregated over trials, participants spent more money for punishing the defection of likable-looking and smiling partners compared to punishing the defection of unlikable-looking and nonsmiling partners, but only because participants were more likely to cooperate with likable-looking and smiling partners, which provided the participants with more opportunities for moralistic punishment. When expressed as a conditional probability, moralistic punishment did not differ as a function of the partners' facial likability. Smiling had no effect on the probability of moralistic punishment, but punishment was milder for smiling in comparison to nonsmiling partners.

Keywords: cooperation; trust; moralistic punishment; facial trustworthiness; facial expression

Objectivity, impartiality and equality are recognized as core principles of justice in many moral (religious, philosophical and legal) systems (DeScioli & Kurzban, 2009). Ideally, decisions about punishment should reflect the nature of the norm-violating behavior, but should be blind to the transgressor's appearance. However, there is reason to assume that the punishment of interaction partners in social dilemmas may be affected by the facial appearance of the partners. One hypothesis is that participants may be more motivated to cooperate with likable-looking and smiling partners and may use punishment to enforce the reciprocation of this cooperation. Another hypothesis is that the defection of a likable-looking or smiling partner violates an appearance-based expectation (Engell, Haxby, & Todorov, 2007; Todorov, 2008; Willis & Todorov, 2006), which may lead to particularly strong feelings of disappointment and anger (Wilson & Eckel, 2006). In consequence, people may impose comparatively more punishment on partners who fail to live up to positive appearance-based expectations. The present study serves to test these hypotheses using a prisoner's dilemma game with a punishment option.

Preliminary evidence for an enhanced punishment of appearance-incongruent behaviors has been obtained in the trust game, but this evidence is difficult to interpret. In the trust game, a truster receives a certain amount of money. The truster can keep the money or send part of it to a trustee. The amount sent is tripled by the experimenter. Then the trustee decides how much of the total amount is sent back to the truster. Wilson and Eckel (2006) observed that attractive trustees received more money (i.e., were trusted more) than unattractive trustees. However, the trustees send less money back to attractive trusters than to unattractive trusters, which represents a "beauty penalty". A possible explanation for this finding is that the trustees expected attractive trusters to send more money than unattractive trusters, and punished the attractive trusters for not living up to these expectations. Although failing to send money back to the truster can thus be interpreted as a form of punishment, the decision to keep most of the money could also be motivated by the selfishness of the trustee (securing a high payoff).

This ambiguity is abolished when the punishment is costly, that is, when participants have to sacrifice resources to punish their partners. This is the case in the ultimatum game, in which participants have to sacrifice their payoff to interfere with a partner's unfair decision. In this game, a proposer receives a certain amount of money, and decides how to divide this sum between him- or herself and a responder. The responder either accepts or rejects this proposal. If the proposal is rejected, neither player receives any money. Although the rational strategy for the responder would be to accept any amount of money,

unfair offers are often rejected. Solnick and Schweitzer (1999) found that attractive responders were offered more than unattractive responders, but attractive proposers had to make higher offers to avoid rejection than unattractive proposers. These results support the idea that appearance-incongruent behaviors receive more punishment. At first glance, not accepting an offer in the ultimatum game can be interpreted as evidence of moralistic punishment. However, note that Solnick and Schweitzer (1999) did not measure the participants' responses to unfair offers, but instead required them to indicate the minimum acceptance level before learning about the proposers' offers. It is therefore possible to speculate that this measure may be more strongly determined by the participants' expectations about the proposers' decisions than by the motivation to punish unfair offers.

Conflicting results have been obtained as well. Ma, Hu, Jiang, and Meng (2015) reported that their participants were less, not more, likely to reject unfair offers from attractive in comparison to unattractive proposers in an ultimatum game. Mussel and colleagues (Mussel, Grritz, & Hewig, 2013; Mussel, Hewig, Allen, Coles, & Miltner, 2014) reported that unfair offers were more likely to be accepted from smiling proposers than from proposers with neutral or angry facial expressions, which was interpreted as suggesting that a smiling facial expression led to a less negative impression of the defector, which in turn resulted in less (not more) punishment. Against the backdrop of these inconsistent findings, the hypothesis that appearance-incongruent defection receives more punishment has to be tested in further experiments before it should be accepted or rejected.

Measuring cooperation and punishment

The present study was designed to examine the effects of facial trustworthiness cues on punishment in a simultaneous prisoner's dilemma game with a subsequent punishment option. In Stage 1, participants played the game with partners who looked likable or unlikable (Experiments 1 and 3) or had a smiling or neutral facial expression (Experiment 2). Both players received a moderate reward (10 cents) for mutual cooperation, and both received nothing in case of mutual defection (0 cents). However, when one player cooperated and the other player defected, unilateral defection was associated with the best payoff (20 cents), while unilateral cooperation led to the worst payoff (−10 cents). Thus, there was an incentive to defect: Independent of what the other player did, each player was better off when refusing cooperation than when cooperating. Nevertheless, both players were better off when both decided to cooperate than when both decided to

defect. In Stage 2, a costly punishment option was given: Participants could invest 0 to 9 cents to punish their partner. For each cent invested, 10 cents were subtracted from the partner's account balance.

As a first step, we analyzed the aggregated amount of money spent to punish the partners in each condition of the 2 (facial appearance: unlikable vs. likable or neutral vs. smiling) by 2 (partner behavior: defection vs. cooperation) design. Analyzing the aggregated amount of money spent to punish the partners is interesting in its own right, but it is important to note that this analysis is incomplete because the punishment of defection likely also depends on the participants' own willingness to cooperate in the first place. If participants have decided to cooperate in Stage 1 of the game, they may be more likely to punish defection than when they have defected themselves. To disentangle the effects of facial likability and one's own willingness to cooperate and to punish, we used a multinomial processing tree model. Historically, these models have been used in genetic analyses to infer the underlying genotypic information from the observable phenotypic category frequencies (Bernstein, 1925). Nowadays such models are commonly used in cognitive psychology to infer latent cognitive states from observable behavior (Erdfelder et al., 2009; Riefer & Batchelder, 1988) or to distinguish between different strategies in decision making (Hilbig, Erdfelder, & Pohl, 2010; Klauer, Stahl, & Erdfelder, 2007; Moshagen, Hilbig, & Musch, 2011). Several computer programs (Moshagen, 2010; Stahl & Klauer, 2007) are available for estimating parameters from observable behavior categories and for performing statistical tests on these parameters.

Here we used the multinomial model shown in Figure 1. The model has five parameters (C , $P_{\text{Moralistic}}$, $P_{\text{Hypocritical}}$, $P_{\text{Irrational}}$ and $P_{\text{Antisocial}}$) and six independent data categories. The rounded rectangles on the left side represent the different types of partners that were encountered during the experiment (defectors and cooperators). Parameter C represents the probability of a participant's willingness to cooperate with either a defector or a cooperator in Stage 1 of the game. The complementary probability $1-C$ represents the probability of a participant's defection in Stage 1 of the game. Note that the willingness to cooperate (that is captured in parameter C) is necessarily independent of whether the partner is a cooperator or a defector because the participant does not know whether the partner will cooperate or defect until this information is revealed after both partners have made their decisions.

The model distinguishes between four different types of punishment. Accordingly, there are four conditional probabilities representing punishment at Stage 2 of the game that depend both on the type of partner (defector or cooperator) as well as on the participant's willingness to cooperate at Stage 1 of the game. Each of these parameters (P_i) represents the conditional probability that a participant decides to punish the other player while the complementary probability ($1-P_i$) represents the probability that a participant decides not to punish the other player.

The following types of punishment were distinguished. (1) Moralistic punishment represented by parameter $P_{\text{Moralistic}}$ is defined as the punishment of the unilateral defection of the partner. It is important that this type of punishment occurs conditional upon a participant's decision to cooperate with the partner in Stage 1 of the game. Moralistic punishment is the most commonly observed type of punishment that is analyzed in social-dilemma games, and is believed to support the enforcing of the norm of cooperation (e.g., Carpenter, 2007; Falk, Fehr, & Fischbacher, 2005). (2) If both the participant and the partner defect, the participant may punish the partner's defection with probability $P_{\text{Hypocritical}}$. One may suspect that this probability should be zero because punishing the partner in this situation would be hypocritical. However, it is well established that defection is punished by persons who do not cooperate themselves, albeit this type of punishment occurs with a smaller probability than moralistic punishment (e.g., Carpenter, 2007; Falk et al., 2005). In line with the literature (Carpenter, 2007), we refer to this type of punishment as "hypocritical" because it is hard to reconcile this behavior with common notions of fairness (Falk et al., 2005). According to Falk et al. (2005), this type of punishment may be motivated by spitefulness (i.e., by the desire to maximize the differences between one's own payoff and that of the other players). (3) Parameter $P_{\text{Irrational}}$ represents the probability that a cooperative participant punishes a cooperative partner. This type of punishment is typically observed with only a very low probability, and it seems difficult to rationally explain this behavior with any theory, which is why we refer to this type of punishment as "irrational". (4) Parameter $P_{\text{Antisocial}}$ represents the probability that an uncooperative participant punishes a cooperative partner. Although the occurrence of antisocial punishment may also seem counterintuitive at first sight, it has been reliably observed across many societies (Herrmann, Thöni, & Gächter, 2008). Cinyabuguma, Page, and Putterman (2006) estimated that about 20% of all punishment decisions fall into this category. Given that unilateral defection is often punished, antisocial punishment may be motivated by the desire to proactively retaliate the (anticipated) punishment of the

cooperator. By representing the punishment parameters as conditional probabilities, the model is capable of obtaining purer measures of the effects of facial likability on punishment that are not confounded by the overall levels of cooperation.

Hypotheses

The following four hypotheses were tested.

(1) Our main hypothesis (Hypothesis 1) was that, aggregated over trials, participants spend more money to punish the appearance-incongruent defection of likable-looking or smiling partners than to punish the appearance-congruent defection of unlikable-looking or nonsmiling partners. As will be explained below, different processes may underlie such an observable behavior pattern. It was therefore necessary to test three more fine-grained hypotheses to gain a deeper understanding of the relationship between facial appearance and punishment.

(2) The aggregated amount of money spent to punish defection depends not only on the partner's behavior (it should be higher when the other player defects than when the other player cooperates), but also on the participant's own behavior in Stage 1 of the game. Cooperative participants should be more concerned with the partner's defection than defecting participants. The moralistic punishment of a partner's unilateral defection should be higher than the hypocritical punishment after mutual defection. In consequence, participants may impose a higher aggregated amount of punishment on likable-looking partners simply because they cooperate more with likable-looking partners, which provides them with more opportunity for moralistic punishment. This possibility can be tested by examining whether the model's cooperation parameter C is higher for likable-looking (or smiling) faces than for unlikable-looking (or nonsmiling) faces (Hypothesis 2a) and whether the model's punishment parameter $P_{\text{Moralistic}}$ is higher than the model's punishment parameter $P_{\text{Hypocritical}}$ (Hypothesis 2b).

(3) The defection of a likable-looking or smiling partner may be a particularly tough emotional experience for cooperative participants. This may motivate them to punish the unilateral defection of their partners. If this hypothesis is correct, the model's parameter $P_{\text{Moralistic}}$ should be higher for likable-looking (or smiling) faces than for unlikable-looking (or nonsmiling) faces (Hypothesis 3).

(4) When appearance-incongruent unilateral defection is perceived as more negative than appearance-congruent unilateral defection, participants may choose harsher

punishment for appearance-incongruent defection (i.e., spend more money to subtract higher amounts of money from their partners' accounts; Hypothesis 4).

Note that any subset of the mechanisms specified in Hypotheses 2 to 4 may cause an increase in the aggregated amount of money spent to punish the defection of likable-looking and smiling partners as opposed to unlikable-looking and nonsmiling partners (Hypothesis 1). Therefore, all of these hypotheses have to be tested to get a complete picture of the relationship between facial appearance and punishment.

Experiment 1

Method

Participants

The sample consisted of 80 Heinrich Heine University students (56 female) aged between 17 and 38 (M of age = 23; SD of age = 4) who participated for course credit or monetary compensation. Mean justice sensitivity (Schmitt, Gollwitzer, Maes, & Arbach, 2005) of the sample was $M = 2.85$ ($SD = 0.63$).

Material

In a norming study (see Bell, Buchner, Kroneisen, & Giang, 2012), likability ratings were obtained for 111 pictures of faces (320 × 240 pixels) of middle-aged females¹ taken from the Center for Vital Longevity Face Database (Minear & Park, 2004). The likability rating scale ranged from 1 (not likable at all) to 6 (extremely likable). Twenty faces with high facial likability ($M = 3.79$; $SD = 0.25$) and 20 faces with low facial likability ($M = 1.96$; $SD = 0.27$) were selected for the experiment.² For each participant, 10 randomly selected likable faces were used as partners who would cooperate with the participant, whereas the remaining 10 faces represented partners who would defect (for details, see the Procedure section). The same was true for the unlikable-looking faces.

¹ We used only female faces because it is well known that female faces are experienced as being somewhat more trustworthy than male faces (Kroneisen & Bell, 2013), and we did not want facial gender to interfere with the facial likability manipulation.

² Please contact the corresponding author to obtain the complete list of stimuli used in the present experiments.

Procedure

Participants played a prisoner's dilemma game with a punishment option. At the start of the game, participants received 400 cents that they could invest. Participants knew from the beginning that they played for real money, and that their account balance was paid to them at the end of the experiment. The first stage of the game was similar to that used in previous studies on the cognitive underpinnings of human cooperation (Bell et al., 2012; Bell, Buchner, & Musch, 2010; Bell, Giang, Mund, & Buchner, 2013; Bell et al., 2016; Giang, Bell, & Buchner, 2012). In each trial, participants saw a silhouette representing themselves on the left side of the screen and the face of their partner on the right side of the screen (Figure 2). The partner's account balance was shown above the face, and was identical to the participant's account balance, +/- a random amount of 0 - 10 cents. The participant decided whether to cooperate or to defect by pressing a button labeled "I cooperate" or "I do not cooperate", respectively. As soon as the participant made a choice, his or her choice and the partner's choice were simultaneously presented on screen for 750 milliseconds. Both investments were represented by arrows that moved towards the center of the screen (within 750 milliseconds). If a player chose to cooperate, he or she invested 30 cents into a joint business with the other player. If a player chose to defect, he or she invested 0 cent. After 750 milliseconds the sum of investments was presented at the center of the screen for 750 milliseconds. A bonus of $\frac{1}{3}$ of the sum of investments was computed and shown for 750 milliseconds. Then the total sum (investment plus bonus) was displayed on the screen. After 750 milliseconds, the total sum was split between the two players. Each player received half of the total sum, regardless of what he or she had invested. Thus, there were three possible outcomes: (1) Both players cooperated, resulting in a gain of 10 cent for each player; (2) one player cooperated and the other defected, resulting in a loss of 10 cent for the cooperative player and a gain of 20 cent for the defector; (3) both players defected, resulting in a payoff of 0 cent for both players. The payoffs were represented by two arrows moving from the center of the screen to the players (within 750 milliseconds). After 1,500 milliseconds, the gain or loss of each player and their updated account balances were shown beneath the participant's silhouette and the partner's face, respectively.

After 750 milliseconds, participants were informed that they now had the option to punish the partner. They could spend between 0 and 9 cents on the partner's punishment. For each cent they invested, 10 cents were subtracted from their partner's account. The

participants typed the amount of money they wanted to spend into a text field and clicked a button labeled “Punishment”. The participant’s choice was then confirmed and the corresponding amount of money was subtracted from the partner’s account. At the same time, the partner’s choice was shown. Cooperative partners invested a random amount varying between 1 and 9 cents to punish participants who defected. Partners who defected themselves did not punish.³ Finally, updated account balances were shown beneath the participant’s silhouette and the partner’s face, respectively. All of the aforementioned information stayed on screen until participants clicked a “Continue” button which initiated the next trial. The participants finished with an average account balance of 167 cents ($SD = 123$). Subsequent to the prisoner’s dilemma game, participants filled out the justice sensitivity questionnaire by Schmitt, Gollwitzer, Maes, and Arbach (2005).

Design

A 2×2 design was used with facial likability (unlikable vs. likable) and partner behavior (defection vs. cooperation) as within-subject factors. The participants’ willingness to cooperate and to punish were analyzed using the multinomial model explained above. In the multinomial analysis, it was possible to detect even very small effects of size $w = 0.06$ (Cohen, 1988), with $\alpha = .05$, $1-\beta = .95$, 40 punishment decisions and $N = 80$ participants. This sensitivity analysis was calculated using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007).

Results

Aggregated amount of money spent to punish the partners

Seventy-two of the 80 participants (90%) punished the partner’s defection at least once. Moreover, 37 of 80 participants (46%) also punished cooperative behavior at least once. The aggregated amount of punishment in each condition (summed across trials and averaged across participants) is shown in Table 1. As expected, there was a significant main effect of partner behavior, $F(1,79) = 87.93$, $p < .001$, $\eta_p^2 = .527$: Participants spent more money punishing defectors than cooperators. Furthermore, there was a significant main effect of facial likability, $F(1,79) = 7.55$, $p = .007$, $\eta_p^2 = .087$. Participants spent more

³ The partners were programmed to show only moralistic punishment because it is well established that this type of punishment is the most common type of punishment experienced by players when interacting with real human partners (Carpenter, 2007; Falk et al., 2005). This is confirmed by the results of the present series of experiments, in which participants used the punishment option primarily for the moralistic punishment of their partners.

money punishing likable-looking partners than unlikable-looking partners. The interaction was also significant, $F(1,79) = 13.40$, $p < .001$, $\eta_p^2 = .145$: Participants spent more money punishing defectors with a likable appearance than defectors with an unlikable appearance, $F(1,79) = 12.19$, $p < .001$, $\eta_p^2 = .134$, and spent more money punishing cooperators with an unlikable appearance than cooperators with a likable appearance, $F(1,79) = 4.69$, $p = .033$, $\eta_p^2 = .056$. This confirms Hypothesis 1 that participants spend more money to punish appearance-incongruent in comparison to appearance-congruent behaviors. However, the aggregated amount of money spent for punishment is influenced not only by the factors that were manipulated (facial likability and partner behavior), but also by the participants' own cooperation or defection in Stage 1 of the game. Therefore, it is necessary to disentangle the effects of facial appearance on cooperation and punishment in the multinomial analysis.

Model-based analysis

Two sets of the processing trees in Figure 1 were used for the multinomial model-based analysis, one for likable-looking and one for unlikable-looking partners. Model fits and statistical tests were performed using multiTree (Moshagen, 2010). The observed response frequencies for all experiments are reported in the Appendix. The base model described above provided a good fit to the data, $G^2(2) = 1.01$, $p = .604$. Parameter estimates are shown in Figure 3.

The willingness to cooperate (captured by parameter C) was higher for interactions with likable-looking partners than for interactions with unlikable-looking partners, $\Delta G^2(1) = 202.92$, $p < .001$, $w = 0.252$. This provides participants with more opportunity for moralistic punishment. As expected, moralistic punishment was more prevalent than hypocritical punishment, $\Delta G^2(2) = 334.01$, $p < .001$, $w = 0.323$. Irrational punishment occurred with the smallest probability and was less common than antisocial punishment, $\Delta G^2(2) = 46.16$, $p < .001$, $w = 0.120$.

The probability of moralistic punishment did not differ between likable-looking and unlikable-looking faces, $\Delta G^2(1) < 0.01$, $p = .967$, $w < 0.001$, disconfirming Hypothesis 3. Likewise, facial likability had no effect on the probabilities of hypocritical punishment, $\Delta G^2(1) = 2.30$, $p = .130$, $w = 0.027$, irrational punishment, $\Delta G^2(1) = 0.46$, $p = .495$, $w = 0.012$, and antisocial punishment, $\Delta G^2(1) = 0.04$, $p = .834$, $w = 0.004$, either.

Harshness of moralistic punishment

There is one further hypothesis that remains to be tested. When the participants decide to punish their partners moralistically, do they invest more money to impose a higher punishment on likable-looking in comparison to unlikable-looking partners? To obtain a measure of the harshness of punishment, the aggregated amount of money spent on moralistic punishment was divided by the number of trials in which this type of punishment was applied. Note that this analysis is necessarily restricted to the subsample of participants who showed moralistic punishment at least once. The harshness of moralistic punishment was similar for likable-looking ($M = 4.24$ cents; $SE = 0.26$) and unlikable-looking ($M = 4.12$ cents; $SE = 0.25$) faces, and did not differ significantly, $F(1,65) = 0.42$, $p = .519$, $\eta_p^2 = .006$.

Discussion

Aggregated over trials, participants spent indeed more money punishing the defection of likable-looking partners than punishing the defection of unlikable-looking partners (confirming Hypothesis 1). However, an analysis of the processes underlying this observable behavior showed that it was solely due to the fact that participants cooperated more with likable-looking partners than with unlikable-looking partners, providing more occasions for the moralistic punishment after unilateral defection by the partners (confirming Hypothesis 2a), which occurred with a higher probability than the hypocritical punishment after mutual defection (confirming Hypothesis 2b). When moralistic punishment was expressed as a conditional probability, there was no evidence of a direct effect of facial likability on the probability of moralistic punishment after unilateral defection (disconfirming Hypothesis 3; the same was true for all other types of punishment). The harshness of moralistic punishment was also not affected by facial likability of the defectors (disconfirming Hypothesis 4). In other words, likable-looking opponents received as much moralistic punishment for unilateral defection as unlikable-looking opponents. These findings clearly disconfirm the idea that appearance-incongruent defection receives disproportionately more or harsher moralistic punishment than appearance-congruent defection. Instead, Experiment 1 suggests that facial likability has indirect rather than direct effects on punishment: Participants are more willing to cooperate with likable-looking opponents than with unlikable-looking opponents, which simply provides them with more opportunity for moralistic punishment if the expectation of mutual cooperation is disappointed.

Experiment 2

Experiment 2 served as a conceptual replication of Experiment 1. We used smiling and nonsmiling faces instead of likable-looking and unlikable-looking faces. As in Experiment 1, these faces were presented before the participants played the game to stimulate positive or negative expectations that could be confirmed or disconfirmed by the partners' cooperation and defection. It was an open question whether there would be a more direct link between smiling and moralistic punishment. Two conflicting hypotheses can be put forward. First, it seems possible that smiling defectors receive more moralistic punishment than neutral-looking defectors. Facial likability and facial attractiveness are used to draw inferences about character traits, but they are stable properties of the faces, and therefore cannot be used to signal specific behavioral intentions. A smiling facial expression, in contrast, may be used to express an invitation to engage into a cooperative interaction and to communicate prosociality (Eckel & Wilson, 2003). As an example, an employer may require the employees to smile at customers to establish the trust that is necessary for engaging in cooperative relationships. Even if the probability or harshness of the moralistic punishment is not increased for behavior that is incongruent with the generally likable facial appearance of a partner (as shown in Experiment 1), there is still the possibility that the false signaling of cooperative intentions via smiling elicits particularly strong negative emotions. This could cause participants to punish the unilateral defection of a smiling partner with a particularly high probability or particularly harshly or both (Brown & Moore, 2002). Experiment 2 served to test this possibility.

However, the opposite prediction (that smiling faces are treated more leniently than nonsmiling faces) can be justified as well. Smiling can be traced back to the bared-teeth display in primates, which is often considered to be a signal of formal subordination, submission, and appeasement (Goldenthal, Johnston, & Kraut, 1981; Lockard, Fahrenbruch, Smith, & Morgan, 1977; Preuschoft, 1992). Therefore, it has been postulated that smiling "is an appeasing device used to prevent or ameliorate intraspecific aggression" (Mackey, 1976, p. 125). If this hypothesis is true, then we might expect to find more lenient treatment of smiling defectors relative to neutral-looking defectors. The same hypothesis can be based on the more general hypothesis that smiling may induce positive emotions, which may, in turn, decrease punishment.

Method

Participants

The sample consisted of 80 Heinrich Heine University students (58 female) aged between 19 and 45 ($M = 24$; $SD = 4$). One dataset had to be excluded because the person had participated twice. Mean justice sensitivity of the sample was $M = 2.77$ ($SD = 0.63$).

Material, Procedure and Design

Material, procedure and design were identical to those of Experiment 1, with the following exceptions. Color photographs (320 × 240 pixels) of 40 middle-aged women were selected from the Center for Vital Longevity Face Database (Minear & Park, 2004). Two photographs of each person were selected, one showing the person with a smiling expression, and one showing the person with a neutral expression. In the prisoners' dilemma game, participants saw the faces of 40 persons, 10 cooperators with a smiling expression, 10 defectors with a smiling expression, 10 cooperators with a neutral expression, and 10 defectors with a neutral expression. The faces were randomly assigned to conditions. A 2 × 2 design was used with facial expression (neutral vs. smiling) and partner behavior (defection vs. cooperation) as within-subject factors.⁴ The participants finished with an average account balance of 155 cents ($SD = 126$).

Results

Aggregated amount of money spent to punish the partners

Seventy-three of 80 participants (91%) punished defection at least once. Moreover, 38 of 80 (48%) participants punished cooperative behavior as well. The aggregated amount of money spent to punish the partners in each condition of the 2 × 2 design (summed across trials and averaged across participants) is reported in Table 1. As in Experiment 1, participants spent more money punishing defectors than cooperators, $F(1,79) = 59.10$, $p < .001$, $\eta_p^2 = .428$. The main effect of facial expression was not significant, $F(1,79) = 1.87$, $p = .176$, $\eta_p^2 = .023$, but there was a significant interaction between partner behavior and facial expression, $F(1,79) = 12.78$, $p < .001$, $\eta_p^2 = .139$. Participants spent more money

⁴ Due to a programming error, 15 of the 80 participants (who pressed the return-button on the keyboard instead of clicking on the on-screen "Punishment" button) saw a blank trial, after which the experiment continued as described. In the present analyses, the data of these 15 participants are included, but the results and the statistical conclusions reported below do not change if these data are removed from the analyses.

punishing smiling defectors than punishing defectors with a neutral facial expression, $F(1,79) = 7.50$, $p = .008$, $\eta_p^2 = .087$, and they spent more money punishing nonsmiling cooperators than punishing smiling cooperators, $F(1,79) = 6.79$, $p = .011$, $\eta_p^2 = .079$.

Model-based analysis

The multinomial model depicted in Figure 1 was used to disentangle the effects of facial expression on cooperation and punishment. Two sets of the processing trees were needed, one for faces with a neutral expression, and one for faces with a smiling expression. The base model provided a good fit to the data, $G^2(2) = 0.28$, $p = .869$.

Parallel to Experiment 1, the willingness to cooperate (reflected in the model's parameter C) was higher in the smiling condition than in the neutral condition, $\Delta G^2(1) = 105.47$, $p < .001$, $w = 0.182$ (left panel of Figure 4). This provides participants with more opportunity for moralistic punishment, which again occurred with a higher probability than all other types of punishment. The parameters representing the conditional probabilities that participants punished the other player are shown in the right panel of Figure 4. Descriptively, the pattern looks very similar to that obtained in Experiment 1. Moralistic punishment occurred with a higher probability than hypocritical punishment, $\Delta G^2(2) = 347.23$, $p < .001$, $w = 0.329$. Irrational punishment occurred with the smallest probability, and was less common than antisocial punishment, $\Delta G^2(2) = 93.29$, $p < .001$, $w = 0.171$.

Parallel to Experiment 1, the probability of moralistic punishment did not differ as a function of facial expression, $\Delta G^2(1) = 0.31$, $p = .575$, $w = 0.010$. The same was true for hypocritical punishment, $\Delta G^2(1) = 0.85$, $p = .358$, $w = 0.016$, irrational punishment, $\Delta G^2(1) = 2.29$, $p = .131$, $w = 0.027$, and antisocial punishment, $\Delta G^2(1) = 0.30$, $p = .582$, $w = 0.010$.

Harshness of moralistic punishment

The moralistic punishment of smiling partners ($M = 4.01$ cents; $SE = 0.30$) was less harsh than the moralistic punishment of nonsmiling partners ($M = 4.31$ cents; $SE = 0.30$), $F(1,66) = 5.77$, $p = .019$, $\eta_p^2 = .080$.

Discussion

The main results of Experiment 1 were replicated in Experiment 2. When analyzing the aggregated amount of money spent to punish the partners, we found that participants

spent more money to punish the defection of smiling partners than to punish the defection of nonsmiling partners. As in Experiment 1, what may at first glance look like an incongruity effect was in fact neither due to a direct effect on the probability with which participants administered moralistic punishment, nor due to a direct effect on the harshness with which participants punished unilateral defection. Instead, participants showed a tendency towards cooperating more with smiling than with nonsmiling partners, which provided more opportunity for engaging in moralistic punishment. As in Experiment 1, participants were more likely to show moralistic punishment compared to hypocritical punishment. These preferences explain why the aggregated amount of money spent to punish the smiling defectors was enhanced.

Despite the fact that the probability of moralistic punishment did not differ between smiling and neutral faces, the analysis of the harshness of moralistic punishment revealed that participants chose somewhat milder moralistic punishments for smiling than for neutral-looking partners. This result is in line with findings in other social dilemma games showing that smiling can decrease punishment (Mussel et al., 2013). Mussel et al. (2014) have shown that unfair offers from smiling proposers in an ultimatum game elicited a less pronounced feedback related negativity (a negative-going neuronal EEG component reflecting the negative evaluation of a feedback) than unfair offers from nonsmiling proposers, which suggests that the smiling of the partner reduced the anger experienced by the participants. These results lend support for the theory that smiling may not only serve to increase trust and to signal cooperative intentions (Eckel & Wilson, 2003), but may also have a function in promoting appeasement and in reducing intraspecific aggression, as predicted by evolutionary theorizing and comparative animal research (Goldenthal et al., 1981; Mackey, 1976). Another possible explanation is that smiling simply elicited positive emotions, which served to reduce the negative response to the partners' unilateral defection.

Experiment 3

Experiment 3 served the purpose of providing an independent replication of the main results of Experiment 1 while also addressing a possible concern about the previous experiments. In both previous experiments, the partners were programmed to punish unilateral defection because it is well established that participants typically experience a high level of moralistic punishment when interacting with real human partners (Carpenter, 2007; Falk et al., 2005). However, it could be argued that the moralistic punishment by the

partners may have created an ad hoc norm that led participants to punish moralistically which they might not have done otherwise. Although it is not quite clear whether and how this might have affected the way in which participants responded to facial appearance, it is of course possible to speculate that this may have influenced the participants' decisions to cooperate and to punish their partners in some yet unknown way. It seemed preferable to eliminate this uncertainty by testing empirically whether the main findings of Experiments 1 and 2 can be replicated when the partners do not serve as an example for how to use the punishment option. Therefore, the punishment option was only available to the participants, but not to the partners, in Experiment 3. In all other aspects of the procedure, Experiment 3 was identical to Experiment 1.

Moralistic punishment serves to enforce cooperation in groups and societies (Fehr & Gächter, 2002). To the extent that participants cooperated only because they anticipated the moralistic punishment of their uncooperative decisions in Experiment 1, we can expect that the willingness to cooperate decreases when participants do not have to fear the punishment of their own uncooperative decisions in Experiment 3. With respect to all other findings, we expected to replicate the main results of Experiment 1. Specifically, we expected to find that moralistic punishment would still occur with a higher probability than all other types of punishment, as in the previous experiments (although the probability of moralistic punishment may slightly decrease in Experiment 3 to the extent that participants followed the example of their partners in the previous experiments). Furthermore, we expected that participants would be more likely to cooperate with likable-looking partners in comparison to unlikable-looking partners. Most importantly, we expected to replicate the finding of Experiment 1 that facial likability has no direct effect on either the probability of moralistic punishment or on the relative strength of moralistic punishment.

Method

Participants

The sample consisted of 80 Heinrich Heine University students (56 female) aged between 17 and 43 (M of age = 24; SD of age = 6).

Material, Procedure and Design

Material, procedure and design were identical to those of Experiment 1, except that the punishment option was only available to the participants, not to the partners. Due to the

lack of punishment by their partners, participants finished with an average account balance of 629 cents ($SD = 104$).

Results

Aggregated amount of money spent to punish the partners

Sixty-four of the 80 participants (80%) punished defection at least once. Moreover, 26 of 80 (33%) participants punished cooperative behavior as well. The aggregated amount of punishment in each condition (summed across trials and averaged across participants) is reported in Table 1. Overall, participants spent less money punishing compared to Experiments 1 and 2. However, they spent more money punishing defectors than cooperators, $F(1,79) = 26.64$, $p < .001$, $\eta_p^2 = .252$, as in the previous experiments. The main effect of facial likability was not significant, $F(1,79) = 2.33$, $p = .131$, $\eta_p^2 = .029$. Also, the interaction between partner behavior and facial likability did not attain significance, $F(1,79) = 2.29$, $p = .134$, $\eta_p^2 = .028$. However, participants showed a tendency towards spending more money to punish the defection of likable-looking partners in comparison to unlikable-looking partners, $F(1,79) = 3.47$, $p = .066$, $\eta_p^2 = .042$.

Model-based analysis

The same base model as in Experiment 1 provided a good fit to the data, $G^2(2) = 1.78$, $p = .411$. As expected, the participants' willingness to cooperate with the partners was somewhat decreased in Experiment 3 in comparison to Experiment 1, probably because they did not have to fear their partners' moralistic punishment in Experiment 3. Nevertheless, the C parameters representing cooperation with likable and unlikable partners were clearly different from zero (left panel of Figure 5). Parallel to Experiments 1 and 2, the willingness to cooperate reflected in the model's parameter C was higher when interacting with likable-looking than when interacting with unlikable-looking partners, $\Delta G^2(1) = 106.00$, $p < .001$, $w = 0.182$.

As in Experiment 1, moralistic punishment occurred with a higher probability than hypocritical punishment, $\Delta G^2(2) = 269.72$, $p < .001$, $w = 0.290$ (right panel of Figure 5). Irrational punishment again occurred with a comparatively small probability, but antisocial punishment occurred with about the same probability as irrational punishment, $\Delta G^2(2) = 0.38$, $p = .836$, $w = 0.011$. This is different from Experiment 1, in which antisocial punishment occurred with a descriptively higher probability than irrational punishment. The

decrease in antisocial punishment in Experiment 3 is to be expected based on the assumption that antisocial punishment is motivated by the desire to retaliate the anticipated moralistic punishment of a cooperative partner (Nikiforakis, 2008). This motivation for antisocial punishment was not present in Experiment 3.

Importantly, and parallel to Experiments 1 and 2, facial likability had no influence on moralistic punishment, $\Delta G^2(1) = 1.06$, $p = .304$, $w = 0.018$. The same was true for hypocritical punishment, $\Delta G^2(1) = 0.43$, $p = .511$, $w = 0.012$, irrational punishment, $\Delta G^2(1) = 0.45$, $p = .504$, $w = 0.012$, and antisocial punishment, $\Delta G^2(1) = 0.15$, $p = .696$, $w = 0.007$.

Harshness of moralistic punishment

As in Experiment 1, and parallel to Experiment 2, the harshness of moralistic punishment did not differ as a function of whether the partners were likable looking ($M = 3.31$ cents; $SE = 0.29$) or unlikable looking ($M = 3.53$ cents; $SE = 0.29$), $F(1,44) = 0.80$, $p = .375$, $\eta_p^2 = .018$.

Discussion

As in Experiments 1 and 2, participants tended to spend more money to punish the defection of likable-looking partners than to punish the defection of unlikable-looking partners. However, this comparison just missed the conventional level of statistical significance in Experiment 3 (and would have been significant when tested as a directional hypothesis). All other findings were very similar to Experiments 1 and 2. As in Experiments 1 and 2, participants were more likely to cooperate with likable-looking compared to unlikable-looking partners while facial likability had neither a direct effect on the probability nor an effect on the harshness of moralistic punishment. Noticeably, moralistic punishment was again the most common type of punishment even though the participants could not follow the example of the partners in Experiment 3.

As expected, the main difference to Experiments 1 and 2 was that the probability of cooperation was decreased in Experiment 3 compared to Experiment 1. This is to be expected given the plausible assumption that the moralistic punishment by the partners had served to increase participants' willingness to cooperate in Experiment 1 and 2 (consistent with Fehr & Gächter, 2002). However, it is noticeable that cooperation did not drop to zero in Experiment 3. Rational choice theory would predict that participants should always choose not to cooperate in Experiment 3 because uncooperative decisions were

associated with a higher payoff. Nevertheless, participants demonstrated a considerable amount of cooperation. This is a typical finding in social dilemma games (Carpenter, 2007; Falk et al., 2005) and suggests that the participants' cooperation in the previous experiments was not solely motivated by their fear of moralistic punishment, but was, to a comparatively large degree, motivated by genuine fairness concerns. The high amount of punishment observed in the present series of experiments leads to a similar conclusion. Given that the participants played with each partner only once, they could not gain anything from punishing their partners. Thus, the rational choice from an egoistic point of view would be not to invest any money to punish uncooperative partners. In contrast to this prediction of rational choice theory, most of the participants punished at least once. Obviously, participants did not only care for their monetary payoff. It seems as if they experienced the punishment of uncooperative decisions as rewarding by itself. These findings are in line with previous findings showing that participants are willing to sacrifice resources to punish uncooperative or unfair behaviors (Carpenter, 2007; Falk et al., 2005; Fehr & Gächter, 2002).

General Discussion

Our starting question was: How is punishment affected by facial likability and facial expression? The answer is somewhat more complex than simply stating that facial likability and smiling enhance punishment. The results show that the aggregated amount of money participants spent to punish defection was affected by the incongruity between the facial appearance and the behavior. Specifically, participants spent a higher aggregated amount of money punishing likable-looking and smiling defectors than unlikable-looking and nonsmiling defectors. Given this result, the next step was to analyze *why* this behavior was observed.

It turned out that the particularly high aggregated amount of money spent punishing appearance-incongruent defection was not due to direct effects of facial appearance on the probability or harshness of moralistic punishment, but was instead solely due to the fact that participants cooperated more with likable-looking and smiling partners and therefore had more occasions for moralistic punishment when their cooperation was not reciprocated. An analysis of the processes underlying the aggregated amount of money spent to punish defection showed that the probability of moralistic punishment was not affected by facial appearance. In addition, likable-looking and smiling defectors were not punished any harsher than unlikable-looking and nonsmiling defectors. The harshness of

moralistic punishment imposed on smiling partners was even less severe than that imposed on nonsmiling partners. Thus, the hypothesis that appearance-incongruent defection leads to particularly likely or harsh moralistic punishment was plausible a priori, but was clearly disconfirmed by the present series of experiments.

Nevertheless, the finding that the aggregated amount of money participants spent to punish the defection of partners who are likable looking or smiling could easily be misinterpreted as evidence for a direct effect of facial appearance on punishment. Thus, this finding illustrates an important point. If a person is likable looking, many others will want to cooperate with this person. If this cooperation is not reciprocated, these others will engage in moralistic punishment. This may cause likable-looking persons to experience a higher amount of moralistic punishment, and a higher pressure to cooperate. Unlikable-looking persons, in contrast, may experience less moralistic punishment because there is a higher probability that others defect against these persons, which gives those others less reason to punish moralistically. Ultimately, these mechanisms could result in a different amount of pressure to cooperate for likable-looking and unlikable-looking persons, even when these effects are not mediated by direct effects of facial likability on the probability or harshness of moralistic punishment in the case of unilateral defection.

A potential concern may be that participants played the social-dilemma game with simulated partners rather than real persons which gave us a high degree of experimental control over the experiences of the participants. Although participants were not informed a priori that they played with computerized partners, it seems possible to speculate that the participants responded in a different way than when interacting with real human partners. A priori, it seemed conceivable that the participants would show purely egoistic behavior when playing for real money against computerized partners. However, this was not the case. Participants were quite willing to sacrifice monetary resources to punish uncooperative behaviors in all three experiments, knowing that they would not gain anything personally from punishing. In fact, the rate of punishment was similar to that observed in experiments in which participants played with real human partners (Falk et al., 2005). Furthermore, participants even chose to cooperate with their partners in Experiment 3 although cooperating was less profitable than defecting and even though participants did not even have to fear their partners' moralistic punishment. Given that participants knew from the beginning that they played for real money, this behavior is irrational from an egoistic point of view, but it is in line with previous observations showing a tendency towards cooperation and norm-conforming behavior in human interactions (Carpenter,

2007; Falk et al., 2005). In sum, the behavior of the participants in the present experiments was very similar to that of participants in studies with real human partners.

Another possible concern is related to the fact that the partners were programmed to engage in moralistic punishment in Experiments 1 and 2. This may have created an implicit norm to punish the defection of others which participants may have followed, creating some uncertainty as to whether, and if so, how that would influence the punishment behavior in general. However, this concern was invalidated by the results of Experiment 3, showing a very similar pattern of punishment behavior as that found in Experiments 1 and 2 but for a situation in which the participants could not follow the example of their partners.

In summary, although plausible a priori, appearance-incongruent defection does not lead to particularly likely, and harsh levels of, moralistic punishment: The probability of moralistically punishing likable-looking and smiling defectors for unilateral defection was just as large as the probability of punishing unlikable-looking and nonsmiling defectors, and likable-looking and smiling defectors were not punished any harsher than unlikable-looking and nonsmiling defectors. Instead, participants simply cooperated more with likable-looking and smiling partners, leading to more occasions for moralistic punishment when their partners did not reciprocate their cooperation.

References

- Bell, R., Buchner, A., Kroneisen, M., & Giang, T. (2012). On the flexibility of social source memory: A test of the emotional incongruity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1512–1529.
- Bell, R., Buchner, A., & Musch, J. (2010). Enhanced old-new recognition and source memory for faces of cooperators and defectors in a social-dilemma game. *Cognition*, *117*, 261–275.
- Bell, R., Giang, T., Mund, I., & Buchner, A. (2013). Memory for reputational trait information: Is social–emotional information processing less flexible in old age? *Psychology and Aging*, *28*, 984–995.
- Bell, R., Sasse, J., Möller, M., Czernochowski, D., Mayr, S., & Buchner, A. (2016). Event-related potentials in response to cheating and cooperation in a social dilemma game. *Psychophysiology*, *53*, 216–228.
- Bernstein, F. (1925). Zusammenfassende Betrachtungen über die erblichen Blutstrukturen des Menschen [Summarizing consideration on the inheritable blood structure of mankind]. *Zeitschrift für Abstammungs- und Vererbungslehre*, *37*, 237–270.
- Brown, W. M., & Moore, C. (2002). Smile asymmetries and reputation as reliable indicators of likelihood to cooperate: An evolutionary analysis. In S. P. Shohov (Ed.), *Advances in Psychology Research* (pp. 59–78). Huntington, New York.
- Carpenter, J. (2007). The demand for punishment. *Journal of Economic Behavior & Organization*, *62*, 522–542.
- Cinyabuguma, M., Page, T., & Putterman, L. (2006). Can second-order punishment deter perverse punishment? *Experimental Economics*, *9*, 265–279.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum, Inc.
- DeScioli, P., & Kurzban, R. (2009). The Alliance Hypothesis for Human Friendship. *PLoS ONE*, *4*, e5802.
- Eckel, C. C., & Wilson, R. K. (2003). The human face of game theory: Trust and reciprocity in sequential games. In E. Ostrom, & J. M. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 245–274). New York: Sage.

- Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala. *Journal of Cognitive Neuroscience*, *19*, 1508–1519.
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. *Zeitschrift für Psychologie /Journal of Psychology*, *217*, 108–124.
- Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, *73*, 2017–2030.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.
- Giang, T., Bell, R., & Buchner, A. (2012). Does facial resemblance enhance cooperation? *PLoS ONE*, *7*, e47809.
- Goldenthal, P., Johnston, R., & Kraut, R. (1981). Smiling, appeasement, and the silent bared-teeth display. *Ethology and Sociobiology*, *2*, 127–133.
- Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science*, *319*, 1362–1367.
- Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2010). One-reason decision making unveiled: A measurement model of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 123–134.
- Klauer, K., Stahl, C., & Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 680–703.
- Kroneisen, M., & Bell, R. (2013). Sex, cheating, and disgust: Enhanced source memory for trait information that violates gender stereotypes. *Memory*, *21*, 167–181.
- Lockard, J. S., Fahrenbruch, C. E., Smith, J. L., & Morgan, C. J. (1977). Smiling and laughter: Different phyletic origins? *Bulletin of the Psychonomic Society*, *10*, 183–186.
- Ma, Q., Hu, Y., Jiang, S., & Meng, L. (2015). The undermining effect of facial attractiveness on brain responses to fairness in the ultimatum game: an ERP study. *Frontiers in Neuroscience*, *9*.
- Mackey, W. (1976). Parameters of the Smile as a Social Signal. *The Journal of Genetic Psychology*, *129*, 125–130.

- Minear, M., & Park, D. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*, 630–633.
- Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42–54.
- Moshagen, M., Hilbig, B. E., & Musch, J. (2011). Defection in the dark? A randomized-response investigation of cooperativeness in social dilemma games. *European Journal of Social Psychology*, *41*, 638–644.
- Mussel, P., Göritz, A. S., & Hewig, J. (2013). The value of a smile: Facial expression affects ultimatum-game responses. *Judgment and Decision Making*, *8*, 381–385.
- Mussel, P., Hewig, J., Allen, J., Coles, M., & Miltner, W. (2014). Smiling faces, sometimes they don't tell the truth: Facial expression in the ultimatum game impacts decision making and event-related potentials. *Psychophysiology*, *51*, 358–363.
- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*, 91–112.
- Preuschoft, S. (1992). “Laughter” and “smile” in barbary macaques (*macaca sylvanus*). *Ethology*, *91*, 220–236.
- Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*, 318–339.
- Schmitt, M., Gollwitzer, M., Maes, J., & Arbach, D. (2005). Justice sensitivity: Assessment and location in the personality space. *European Journal of Psychological Assessment*, *21*, 202–211.
- Solnick, S., & Schweitzer, M. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Processes*, *79*, 199–215.
- Stahl, C., & Klauer, K. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods*, *39*, 267–273.
- Todorov, A. (2008). Evaluating faces on trustworthiness. *Annals of the New York Academy of Sciences*, *1124*, 208–224.
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*, 592–598.
- Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly*, *59*, 189–202.

Figures

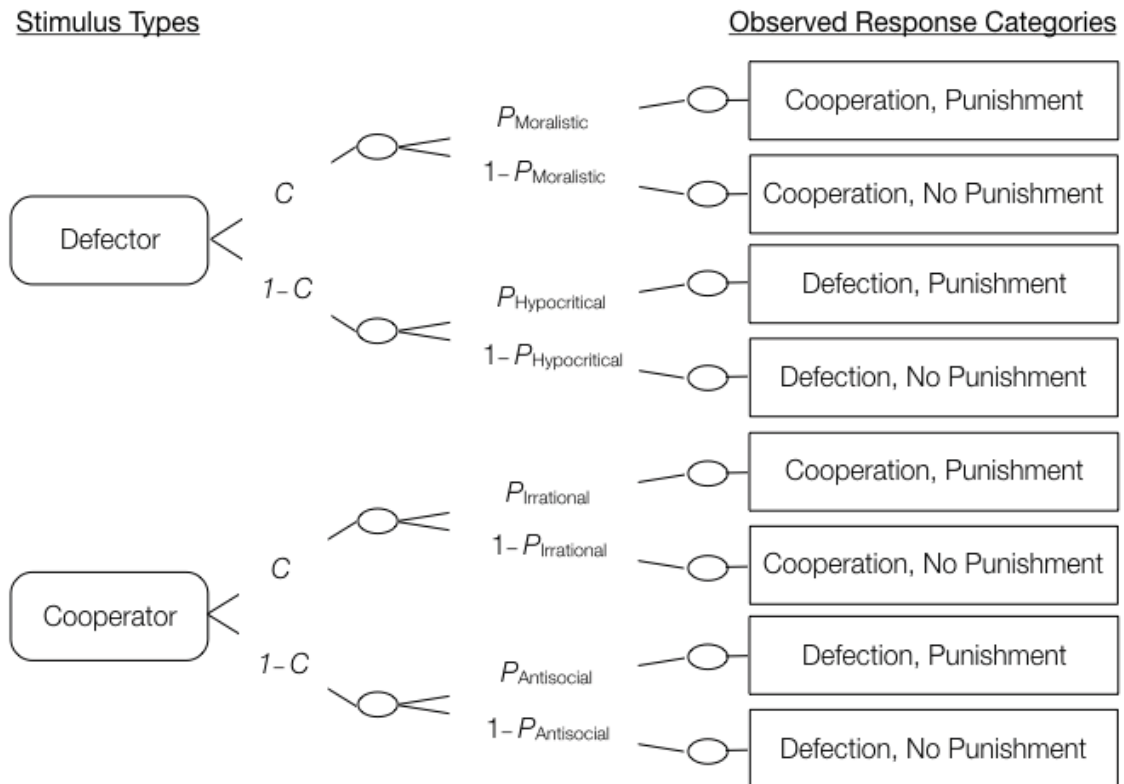


Figure 1. The model of cooperation and punishment described in the text. Rounded rectangles on the left side represent the stimulus persons (uncooperative and cooperative partners), and the rectangles on the right side represent the observable response categories. Parameter C represents the probability that a participant cooperates rather than defects. Parameters P represent the conditional probabilities with which certain types of punishment are administered (moralistic, hypocritical, irrational, and antisocial punishment).

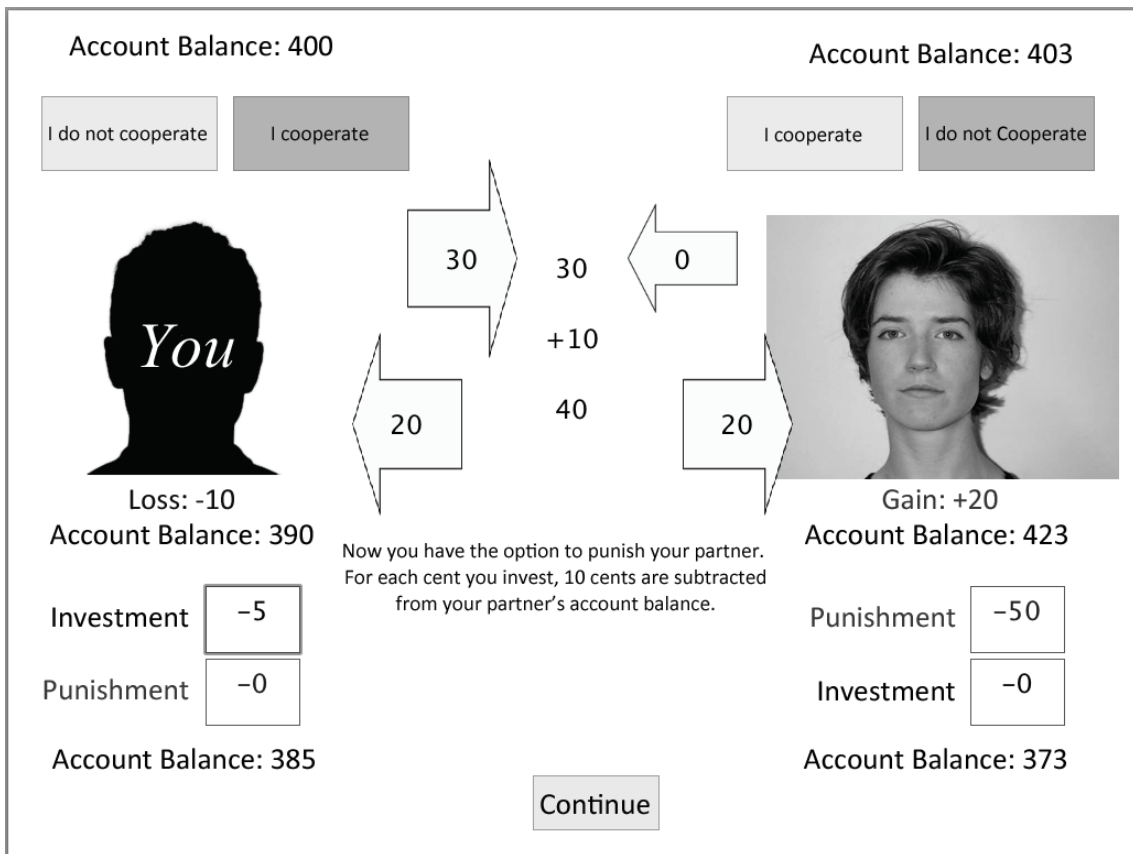


Figure 2. A screenshot of Experiments 1 and 2. The photograph on the right side was taken from the Center for Vital Longevity Face Database (Minear & Park, 2004).

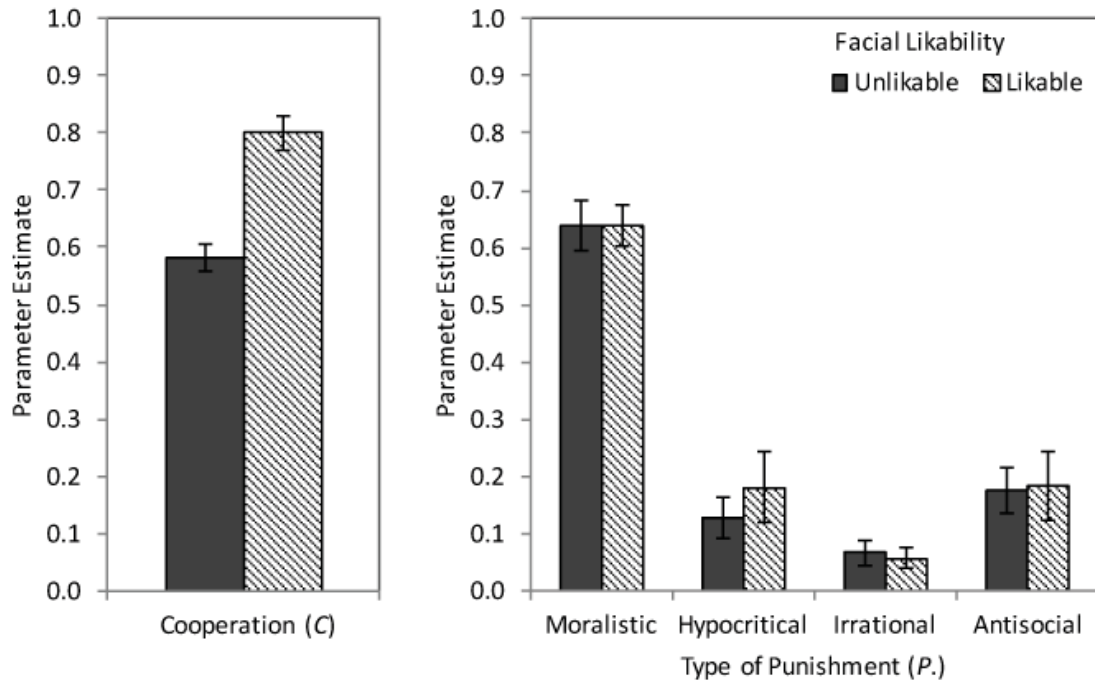


Figure 3. Parameter estimates for Experiment 1. Left panel: Parameter *C* representing the probability of participants' cooperation as a function of facial likability (unlikable vs. likable). Right panel: Punishment as a function of type of punishment (moralistic, hypocritical, irrational and antisocial) and the partner's facial likability (unlikable vs. likable). The error bars represent the .95-confidence intervals.

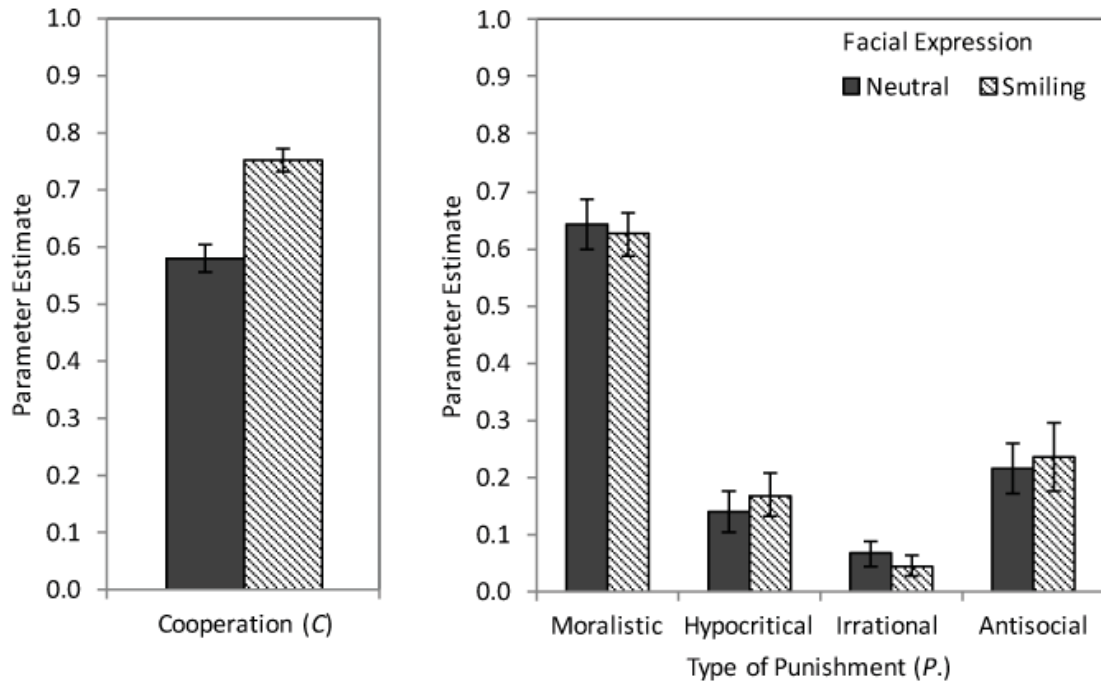


Figure 4. Parameter estimates for Experiment 2. Left panel: Parameter *C* representing the probability of the participants' cooperation as a function of facial expression (neutral vs. smiling). Right panel: Parameters *P*. representing the conditional probabilities of punishment as a function of the type of punishment (moralistic, hypocritical, irrational and antisocial) and facial expression (neutral vs. smiling). The error bars represent the .95 confidence intervals.

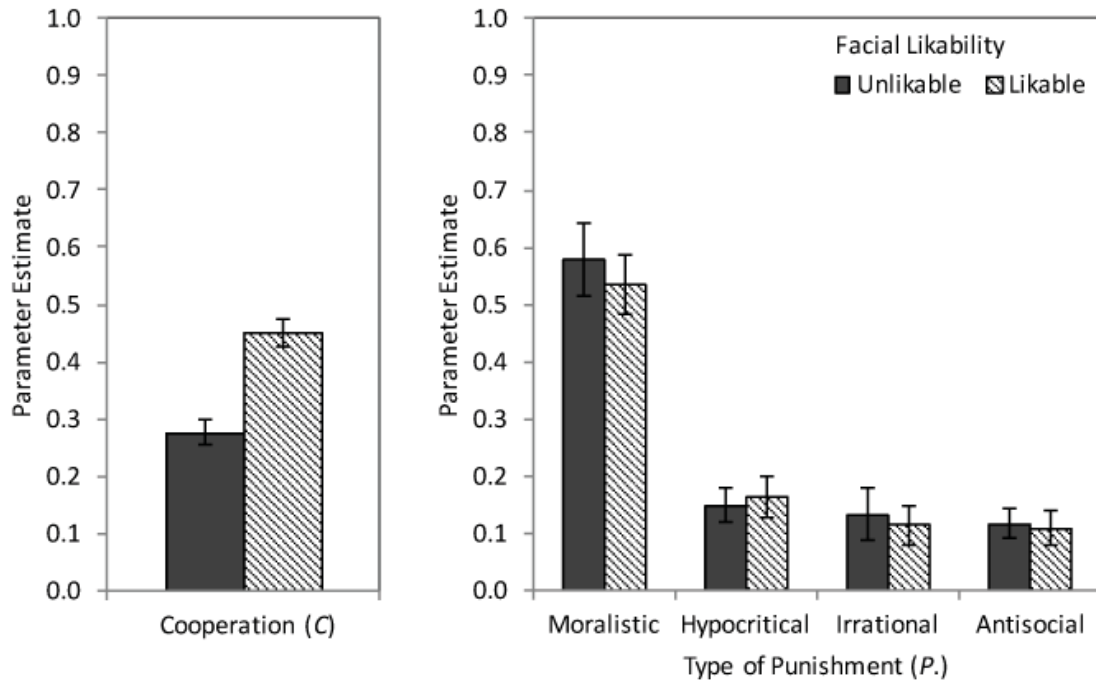


Figure 5. Parameter Estimates for Experiment 3. Left panel: Parameter C representing the probability of the participants' cooperation as a function of facial likability (unlikable vs. likable). Right panel: Conditional probabilities of punishment parameters P . as a function of the type of punishment (moralistic, hypocritical, irrational and antisocial) and the partner's facial likability (unlikable vs. likable). The error bars represent the .95 confidence intervals.

Tables

Table 1: The aggregated amount of money spent to punish the partners in Cent (summarized across trials and averaged across participants) as a function of facial appearance (unlikable vs. likable or neutral vs. smiling) and partner behavior (defection vs. cooperation).

| Facial Appearance | Partner Behavior | | | |
|-------------------|------------------|---------------|-------------|---------------|
| | Defection | | Cooperation | |
| | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) |
| Experiment 1 | | | | |
| Unlikable | 19.15 | (2.40) | 3.94 | (0.99) |
| Likable | 24.26 | (2.33) | 2.65 | (0.79) |
| Experiment 2 | | | | |
| Neutral | 18.80 | (2.27) | 4.45 | (0.89) |
| Smiling | 22.33 | (2.55) | 2.86 | (0.73) |
| Experiment 3 | | | | |
| Unlikable | 9.23 | (1.59) | 3.13 | (0.90) |
| Likable | 10.86 | (1.66) | 2.86 | (0.80) |

Appendix

Observed response frequencies for the participants' punishment or non-punishment as a function of the facial appearance and partner behavior (rows) and the participants' cooperation or defection (columns).

| Stimulus Type | | Participant Behavior | | | |
|-------------------|------------------|----------------------|---------------|------------|---------------|
| | | Cooperation | | Defection | |
| Facial Appearance | Partner Behavior | Punishment | No Punishment | Punishment | No Punishment |
| Experiment 1 | | | | | |
| Unlikable | Defection | 295 | 166 | 43 | 296 |
| | Cooperation | 31 | 439 | 58 | 272 |
| Likable | Defection | 419 | 237 | 26 | 118 |
| | Cooperation | 29 | 129 | 36 | 606 |
| Experiment 2 | | | | | |
| Neutral | Defection | 297 | 166 | 47 | 290 |
| | Cooperation | 31 | 434 | 72 | 263 |
| Smiling | Defection | 378 | 227 | 33 | 162 |
| | Cooperation | 27 | 569 | 48 | 156 |
| Experiment 3 | | | | | |
| Unlikable | Defection | 133 | 97 | 85 | 485 |
| | Cooperation | 28 | 183 | 69 | 520 |
| Likable | Defection | 197 | 171 | 71 | 361 |
| | Cooperation | 40 | 312 | 49 | 399 |

Effects of gender on costly punishment

Laura Mieth
Axel Buchner
Raoul Bell

Heinrich Heine University Düsseldorf

Running Head: Gender and Punishment

Keywords: cooperation; moralistic punishment; social interaction; gender; trust

Corresponding Author:

Laura Mieth

Department of Experimental Psychology

Heinrich Heine University Düsseldorf

40225 Düsseldorf, Germany

E-Mail: laura.mieth@hhu.de

Abstract

Women are often believed to be more cooperative and less egoistic than men. In the present study, we examined whether people punish women for failing to live up to these positive gender stereotypes. Participants played a prisoner's dilemma game with female and male partners who either cooperated or defected. Participants were offered a costly punishment option. They could spend money to decrease the payment of their partners. In Experiment 1, participants spent more money to punish the defection of female in comparison to male partners, but this effect of gender on punishment was indirect rather than direct: Participants were more likely to cooperate with female partners than with male partners, which gave them more opportunity for moralistic punishment. In Experiments 2 and 3, we examined the effects of the participants' own gender on cooperation and punishment of female and male partners. Female participants cooperated more with female partners than with male partners while male participants treated female and male partners equally. We conclude that the effect of facial gender on punishment are indirect rather than direct.

Although women are, relative to men, still disadvantaged in terms of power in many societies, gender stereotypes often paint women in a positive light. For instance, both men and women are more likely to ascribe attributes such as helpful and kind to women than to men, and they are more likely to ascribe attributes such as competitive and dominant to men than to women (Eagly & Mladinic, 1989). These benevolent gender stereotypes imply that women are generally seen as less egoistic and more cooperative than men (although research on actual gender differences in cooperation paints a mixed picture on this issue; see Balliet, Li, Macfarlan, & Van Vugt, 2011).

It seems possible that these positive expectations towards women may translate into increased punishment of behaviors that do not conform to the female role of being selfless and kind. There is already some evidence that selfish and uncooperative behavior is met with particularly strong negative emotions when it violates positive expectations (Wilson & Eckel, 2006). Given that people often expect women to be less driven by self-interest than men (Eagly & Mladinic, 1989), it seems possible that people may enforce different, and stricter, moral standards on women than on men. Specifically, a woman who is selfish and uncooperative may be punished for violating the high social expectations that are part of the female gender role. In this way, a positive view of women's nature may lead to social sanctions against women, and to an increased social pressure to refrain from behaviors that are driven by selfish motives.

In the present study, we tested the hypothesis that participants are more likely to punish the uncooperative behavior of female interaction partners than the uncooperative behavior of male interaction partners. The available literature does not provide definitive evidence on this issue. Preliminary evidence on gender and punishment was obtained in a classic ultimatum-game study by Solnick and Schweitzer (1999). In the ultimatum game, a proposer receives a certain amount of money and has to decide how to split this money between him- or herself and a responder. The responder can either accept the offer, in which case the money is divided as proposed, or reject the offer, in which case none of the players gets any money. Responders often reject offers they perceive to be unfair even at the cost of losing money themselves, which can be interpreted as a form of moralistic punishment. Using this game, Solnick and Schweitzer showed that female proposers are expected to offer higher amounts of money than male proposers. They also demonstrated that offers of female proposers are more likely to be rejected than offers from male proposers. At first glance, these findings seem to support the idea that unfair behavior of women is more likely to be punished than unfair behavior of men. However, it is unclear

whether these findings can be unambiguously interpreted as evidence of moralistic punishment because participants did not *respond* to unfair offers of the proposers, but were instead required to determine the minimal amount of money they would accept *before* the proposers actually made their offer. It seems possible to speculate that the participants' expectations about the future behavior of their partners may play a greater role when determining a minimal acceptance level before playing the game than when retrospectively determining a punishment in response to an unfair behavior. It cannot be taken for granted that these findings generalize to paradigms that measure the retrospective punishment of unfair and egoistic behaviors. It seems therefore necessary to test the hypothesis of a gender-specific increase in punishment in additional studies.

In the present study, participants played a two-stage prisoner's dilemma game with punishment option. In the first stage of the game, participants interacted with partners whose faces were shown on the screen. The participants and their partners decided simultaneously whether they wanted to cooperate or defect. The game was a typical prisoner's dilemma game, which means that mutual cooperation led to a higher payoff than mutual defection, but unilateral defection had the highest payoff while unilateral cooperation resulted in the lowest payoff. Therefore, each player could maximize his or her payoff by choosing to defect, regardless of whether the other player cooperated or defected. However, when both players gave in to this selfish temptation, the outcome for both was worse than when both would have acted cooperatively. In the second stage of the game, a punishment option was introduced: The participant could invest money to punish the partner by subtracting money from the partner's account.

This paradigm was already used to examine the effects of trustworthy facial appearance on punishment in a previous study (Mieth, Bell, & Buchner, 2016). Facial trustworthiness (Todorov, 2008) indeed had an effect on punishment, but the effect was indirect rather than direct. When the amount of punishment was aggregated over trials, participants spent more money punishing uncooperative partners who looked trustworthy than partners who looked untrustworthy. At first glance, this finding could be interpreted as evidence of increased punishment of expectancy-violating behaviors. However, this effect was solely due to the fact that people were more likely to cooperate with trustworthy-looking partners than with untrustworthy-looking partners. It seems clear that people are more motivated to punish their partners' defection after having chosen to cooperate than after having chosen to defect. Therefore, the increased willingness to cooperate with trustworthy-looking partners simply gave the participants more opportunity to moralistically punish the

partners' defection when their own cooperation was not reciprocated. Based on these results, it seems clear that the participants' own behaviors have to be taken into account when measuring the effect of gender on costly punishment.

Measuring punishment

To analyze punishment, we first test whether the aggregated amount of money spent to punish the partners' defection or cooperation varies as a function of gender. However, it is necessary to complement this analysis with one that takes into account the participants' own behaviors because participants may be more likely to punish defection when they have decided to cooperate than when they have decided to defect. To this end, we used the multinomial model shown in Figure 1 (Mieth et al., 2016). Multinomial models are often used in Cognitive Psychology to identify the processes underlying observable behavior (Bayen, Murnane, & Erdfelder, 1996; Bröder & Meiser, 2007; Erdfelder et al., 2009; Klauer, Stahl, & Erdfelder, 2007; Riefer & Batchelder, 1988). The model shown in Figure 1 has already been used successfully to disentangle the effects of facial trustworthiness on cooperation and punishment (Mieth et al., 2016).

In this model, punishment is assumed to vary as a function of both the partners' defection or cooperation as well as the participants' own defection or cooperation. The upper tree shown in Figure 1 represents the participants' behavior in response to defecting partners. Parameter C represents the probability that participants decide to cooperate in Stage 1 of the game. The complementary probability $(1 - C)$ represents the probability that participants decide to defect. Note that the participants' decision to cooperate or defect is necessarily independent of the partners' decision to cooperate or defect as both choices are revealed simultaneously. Therefore, the same probability C represents the probability that participants cooperate with defecting and cooperating partners. Punishment, in contrast, is assumed to differ as a function of the partners' cooperation or defection. Furthermore, the punishment parameters represent conditional probabilities that depend on the participants' own cooperation or defection. Hence, four types of punishment result from the possible combinations of the partners' behaviors (cooperation or defection) and the participants' behaviors (cooperation or defection).

After having cooperated with a defecting partner (with probability C), participants may punish this partner for his or her unilateral defection with probability $P_{\text{Moralistic}}$. This type of moralistic punishment is the most common form of punishment (Carpenter, 2007; Falk, Fehr, & Fischbacher, 2005; Herrmann, Thöni, & Gächter, 2008), and, therefore, the type of

punishment we first think about when we think about costly punishment. Participants can be assumed to have less reason to punish the defection of a partner after having defected themselves (with probability $1 - C$), because doing so would be hypocritical. There is evidence that people do indeed impose hypocritical punishment on their partners (Carpenter, 2007; Falk et al., 2005), but with a smaller probability than moralistic punishment. The probability of hypocritical punishment in this situation is captured in parameter $P_{\text{Hypocritical}}$.

The lower tree in Figure 1 shows the participants' responses to cooperating partners. When participants cooperate with cooperating partners (with probability C), there is little reason to punish the partners. Therefore, we named this type of punishment irrational punishment— $P_{\text{Irrational}}$ —and assumed that it would occur with a very small probability. At first glance, one may also suspect that participants have little reason to impose punishment when having defected against a cooperating partner (with probability $1 - C$), but it is well known that people may impose antisocial punishment on others for unwanted cooperation. The probability of antisocial punishment $P_{\text{Antisocial}}$ can be assumed to be smaller than the probability for moralistic punishment $P_{\text{Moralistic}}$, but antisocial punishment is nevertheless reliably found in many societies (Henrich et al., 2006; Herrmann et al., 2008).

Hypotheses

This model allowed us to test two different hypotheses about why people may impose more punishment on women than on men. To anticipate, such a pattern was indeed observed in Experiments 1 and 2. It may occur either as a direct or as an indirect effect of gender on punishment. The multinomial model illustrates this point. According to the model equations, the probability of punishing the defection of a partner is given by $C \cdot P_{\text{Moralistic}} + (1 - C) \cdot P_{\text{Hypocritical}}$. Let us assume in a hypothetical example that the probability for cooperating with male partners is $C = .50$, the probability of punishing male partners moralistically is $P_{\text{Moralistic}} = .60$, and the probability of punishing male partners hypocritically is $P_{\text{Hypocritical}} = .10$. According to the equation given above, this results in an aggregated probability of $.50 \cdot .60 + .50 \cdot .10 = .35$, with which participants punish the male partners' defection. A direct effect of gender on punishment would be reflected in a higher probability $P_{\text{Moralistic}}$ to punish female partners in comparison to male partners. For instance, if the only difference between female and male partners were that the probability of moralistic punishment of female partners would be $P_{\text{Moralistic}} = .90$, then the defection of female partners would be punished with the aggregated probability of $.50 \cdot .90 + .50 \cdot .10 = .50$.

However, the same finding of a higher aggregated punishment of female in comparison to male partners could result from an indirect effect of gender on punishment. Due to the positive female stereotypes mentioned above, people may tend to cooperate more with female than with male partners. An indirect effect of gender on punishment could be implemented in the model by assuming that (a) the probability that participants cooperate with female partners is higher than the probability that they cooperate with male partners, and that (b) moralistic punishment occurs with a higher probability than hypocritical punishment. For instance, if we assume that people punish female partners with the exact same probabilities as they punish male partners in the example above ($P_{\text{Moralistic}} = .60$, $P_{\text{Hypocritical}} = .10$), but postulate that people cooperate with a particularly high probability of $C = .80$ with female partners, the aggregated probability of punishing the defection of female partners is given by $.80 \cdot .60 + .20 \cdot .10 = .50$, which is higher than the aggregated probability of punishing the defection of male partners given by $.50 \cdot .60 + .50 \cdot .10 = .35$.

This analysis shows that the effect of the partners' gender on punishment may be either direct or indirect. It seems possible that participants impose moralistic punishment with a higher probability on female in comparison to male partners. However, it is also possible that the participants cooperate more with female than with male partners. This fact alone may result in more punishment of the female partners' defection because more cooperation creates more opportunity for moralistic punishment. Finally, it may of course be the case that participants cooperate more with females than with males *and* punish defecting females with a higher probability than they punish defecting males.

To summarize, the present study serves to test the hypothesis that participants are more likely to punish the defection of female partners than the defection of male partners. In addition, we wanted to test whether the effect of gender on punishment (that was indeed observed in Experiments 1 and 2) is direct or indirect or both. A direct effect would imply that participants impose more moralistic punishment on female in comparison to male partners, which implies that the parameter reflecting moralistic punishment $P_{\text{Moralistic}}$ should be higher for female in comparison to male faces. An indirect effect would imply that participants punish the defection of female partners more, but only because female partners are met with more cooperation (reflected by parameter C), which provides participants with more opportunity for moralistic punishment when the partners fail to reciprocate this cooperation.

As yet, we have only discussed the probability of punishment. The picture is incomplete without an analysis of the harshness of punishment imposed on the partners. When the participants decide to punish their partners moralistically, do they invest more money to impose a higher punishment on female in comparison to male partners? This question was analyzed in a supplementary analysis.

Experiment 1

Method

Participants

Eighty students at Heinrich Heine University (61 of whom were female) participated for course credit or monetary compensation. Their ages ranged between 18 and 40 years ($M = 23$, $SD = 4$).

Material

Pictures of 20 female faces and 20 male faces were randomly drawn from a set of 65 pictures of white women between 18 and 29 years and from a set of 65 pictures of white men between 18 and 29 years that were available in the Center for Vital Longevity (CVL) face database (Minear & Park, 2004). We used only frontal-view faces with a neutral facial expression. The pictures were presented at a resolution of 320 × 240 pixels.

Procedure

Participants played a simultaneous prisoner's dilemma game with a punishment option (Mieth et al., 2016). The prisoner's dilemma game was similar to the one used in previous memory studies (Bell, Buchner, Kroneisen, & Giang, 2012; Bell, Buchner, & Musch, 2010; Giang, Bell, & Buchner, 2012). Participants knew from the beginning that they played for real money and that their account balance was payed to them at the end of the experiment. Participants interacted with 20 female and 20 male partners in random order. All participants started with an account balance of 400 cents. In each trial, participants saw a silhouette representing themselves on the left side of the screen and the partner's face on the right side of the screen. They could decide whether they wanted to cooperate or defect by pressing buttons labeled "I cooperate" or "I do not cooperate". Cooperating meant to invest 30 cents into the game while defecting meant investing 0 cents. Half of the partners cooperated, and the other half cheated, with ten trials of each combination of gender (female or male) and behavior (cooperation or defection). In the example shown in

Figure 2, the participant decided to cooperate, and the partner decided to defect. Both choices were displayed simultaneously. The buttons for cooperation and defection were highlighted for 750 milliseconds before the arrows representing both partners' investments moved to the middle of the screen within 750 milliseconds. The arrows stopped for 750 milliseconds. Next, the investments were summed up in the middle of the screen. After 750 milliseconds, a bonus for cooperation ($\frac{1}{3}$ of the total investments) was added and shown for 750 milliseconds. Then the total sum was shown for 750 milliseconds. The total sum was split up between the participant and the partner, and both received half of the total sum regardless of what they had invested. Arrows representing each partner's share moved from the middle of the screen to both sides of the screen within 750 milliseconds. Gains or losses and the updated account balances were shown after 1,500 milliseconds beneath the participant's silhouette and the partner's photograph, respectively.

In this game, mutual cooperation resulted in a moderate gain of 10 cents for both partners. In the example shown in Figure 2, the game resulted in a large gain of 20 cents for the defecting partner who benefitted from the participants' cooperation but did not reciprocate anything in return. The cooperating participant, in contrast, suffered a loss of 10 cents because he had invested more than he got back, due to the partner's lack of cooperation.

Immediately after this interaction, participants could spend between 0 and 9 cents to punish their partners. For each cent they invested, ten cents were subtracted from the partner's account. Participants typed the amount of money they wanted to spend into a text field and confirmed their choice by pressing a button labeled "punishment". The punishment (i.e., the amount subtracted from the partner's account) was shown at the other side of the screen.

Furthermore, the partners had the opportunity to punish the participants. Given that it is well established that moralistic punishment is usually the most frequently experienced type of punishment in interactions with real human partners (Carpenter, 2007; Falk et al., 2005), we decided that participants should experience moralistic punishment within the game. Therefore, cooperating partners imposed moralistic punishment of 10 to 90 cents on defecting participants by spending a random amount between 1 and 9 cents. The influence of the participants' punishment by the partners will be examined in Experiment 3. To anticipate, being punished for defection by cooperating partners increased the participants' cooperation, but did not affect the overall pattern of results.

Finally, the participant's and the partner's updated account balances were shown beneath the silhouette and the partner's face, respectively. Then the participant initiated the next trial by pressing a "continue" button. At the end of the experiment the account balance was payed out to the participants.

Design

The design was a 2×2 design with partner behavior (defection vs. cooperation) and partner gender (female vs. male) as within-subject factors. A sensitivity analysis in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) showed that it was possible to detect small gender differences of the size of $w = 0.06$ (Cohen, 1988) in the multinomial analysis, given $\alpha = .05$, $1 - \beta = .95$, 40 trials and $N = 80$ participants.

Results

Aggregated amount spent to punish female and male partners

Given that participants did not benefit from punishing their partners but had to sacrifice money to do so, the first interesting question is whether they used the punishment option at all. Most of the participants—67 of the 80 participants (84%)—punished their partners' defection at least once. Thirty-six of the 80 participants (45%) even punished their partner for cooperating at least once.

The main question of Experiment 1 was whether punishment depends on the partners' gender. Table 1 shows the average amount of money spent to punish female and male partners, aggregated over trials. Unsurprisingly, defecting partners were punished more than cooperating partners, $F(1,79) = 55.24$, $p < .001$, $\eta_p^2 = .41$. The main effect of partner gender was not significant, $F(1,79) = 0.60$, $p = .439$, $\eta_p^2 < .01$. However, the interaction between partner behavior and partner gender was significant, $F(1,79) = 7.00$, $p = .010$, $\eta_p^2 = .08$: Female partners were punished more for defection than male partners whereas male partners were punished more for cooperation than female partners. At first glance, this pattern of results seems to provide evidence for the hypothesis that behaviors are punished more when they are inconsistent with common gender stereotypes. However, as explained in the Introduction, while this pattern could be due to a direct effect of partner gender on punishment, it may also be due to an indirect effect: If it is true that participants tend to cooperate more with female than with male partners, then the participants also have more opportunity for moralistic punishment. Therefore, we have to account for the

possible effects of the participants' own cooperation on punishment by testing whether facial gender has direct or indirect effects on punishment.

Model-based analysis

To examine the influence of gender on cooperation and punishment, we need two sets of the model trees shown in Figure 1: one set for female partners and one set for male partners. Model fits and parameter estimates were obtained using multiTree (Moshagen, 2010). The model fit the data well, $G^2(2) = 0.09$, $p = .957$.

Parameter C , representing the probability that a participant cooperates rather than defects, is shown in the left panel of Figure 3. Participants were more likely to cooperate with female than with male partners, $\Delta G^2(1) = 23.99$, $p < .001$, $w = 0.09$.

The punishment parameters are shown in the upper panel of Figure 4. As expected, moralistic punishment was more common than hypocritical punishment, $\Delta G^2(2) = 257.49$, $p < .001$, $w = 0.28$, and any other form of punishment. As expected, irrational punishment occurred with the smallest probability, and was less common than antisocial punishment, $\Delta G^2(2) = 114.94$, $p < .001$, $w = 0.19$.

Importantly, partner gender had no direct effect on punishment when the different types of punishment were expressed as conditional probabilities. Participants imposed moralistic punishment, $\Delta G^2(1) = 0.38$, $p = .539$, $w = 0.01$, hypocritical punishment, $\Delta G^2(1) = 0.20$, $p = .652$, $w < 0.01$, irrational punishment, $\Delta G^2(1) = 0.05$, $p = .825$, $w < 0.01$, and antisocial punishment, $\Delta G^2(1) = 0.07$, $p = .790$, $w < 0.01$, on female and male partners with the same probabilities.

Thus, this analysis confirms that partner gender had indirect rather than direct effects on punishment. Participants were more likely to cooperate with female partners than with male partners in Stage 1 of the game, which provided them with more opportunity for moralistic punishment.

Harshness of moralistic punishment

To obtain a measure of the harshness of punishment, the aggregated amount of money spent on moralistic punishment was divided by the number of trials in which this type of punishment occurred. This analysis is necessarily restricted to the data from participants who decided to use moralistic punishment at least once. Moralistic punishment was the only type of punishment that occurred with a high enough probability to be meaningfully analyzed. The harshness of moralistic punishment did not differ as a function of the

partners' gender, $F(1,62) = 0.19$, $p = .668$, $\eta_p^2 < .01$ (see Table 2). This confirms that there is no direct effect of the partners' gender on moralistic punishment.

Discussion

Aggregated over trials, participants spent more money to punish female partners than to punish male partners. At first glance, this result may suggest that people impose an extra punishment on female partners for violating the positive female stereotype of being selfless and kind. However, when analyzed conditionally upon whether cooperation occurred, moralistic punishment did not differ between female and male partners. There was also no direct effect of the partners' gender on the harshness of the moralistic punishment. The effect of facial gender on punishment seemed to be indirect rather than direct: Participants were more willing to cooperate with female partners than with male partners, and this, in turn, provided them with more opportunity for moralistic punishment, which was imposed with a comparatively high probability on both female and male partners in comparison to other types of punishment.

Experiment 2

In Experiment 1, we only examined the effect of the partners' gender on cooperation and punishment, but we left aside the possible influence of the participants' gender. Given that only a small proportion of the participants (19 out of 80) were male in Experiment 1, it was impossible to examine the influence of the participants' gender in this experiment. It is therefore unclear whether the findings would generalize to a sample with more male participants.

Experiment 2 had two aims. The first aim was to replicate the findings of Experiment 1 with new stimulus sets. The second aim was to examine the potential role of the participants' gender. Therefore, we planned to collect data from two equally sized female and male samples. With respect to the female participants, we expected to replicate the findings of Experiment 1 that they tend to cooperate more with female partners and impose an equal amount of moralistic punishment on female and male partners. With respect to the male participants, it was an open question whether or not they would treat female and male partners differently.

By comparing the results of female and male participants, we were also able to examine potential gender differences in cooperation and punishment. If it is true that men are indeed more interested in maximizing profit whereas women care more about the social

aspects of the task (Kahn, Hottes, & Davis, 1971), we may expect, at first glance, that (a) women should show higher levels of cooperation than men, and (b) women should use the costly punishment option more than men. Note, however, that Hypothesis (a) is not applicable to Experiment 2 because participants were moralistically punished for defection by their partners, which implies that they might cooperate in Experiment 2 for the selfish reason of avoiding moralistic punishment. This potential role of the moralistic punishment by the partners on the participants' cooperation is tested in Experiment 3. Hypothesis (b) can be tested because there was no direct incentive for using the costly punishment option in any of the experiments.

Method

Participants

Eighty female and eighty male students at Heinrich Heine University participated for course credit or monetary compensation. Their ages ranged between 18 and 47 years ($M = 22$; $SD = 4$).

Material, Procedure and Design

Material, procedure and design were identical to Experiment 1, with the following exceptions. To examine whether the results would generalize to different sets of faces, 20 female and 20 male faces were randomly sampled from the CVL face database with the restriction that the faces that had been used in Experiment 1 could not be sampled again.

The design was a $2 \times 2 \times 2$ design with partner behavior (defection vs. cooperation), partner gender (female vs. male), and participant gender (female vs. male). Participants' gender was the only between-subjects factor.

Results

Aggregated amount of money spent to punish the partners

As in Experiment 1, a large proportion of participants—75 of the 80 female participants (94%) and 73 of the 80 male participants (91%)—punished the partners' defection at least once. Furthermore, 34 of the 80 female participants (42%) and 29 of the 80 male participants (36%) even punished cooperation at least once.

Table 1 shows the aggregated amount of money spent to punish female and male partners. Participants spent more money to punish defecting partners than to punish cooperating partners, $F(1, 158) = 204.88$, $p < .001$, $\eta_p^2 = .56$. The main effect of partner

gender was not significant, $F(1,158) = 0.04$, $p = .847$, $\eta_p^2 < .01$. The interaction between partner behavior and partner gender was significant, $F(1,158) = 4.83$, $p = .029$, $\eta_p^2 = .03$: Female partners were punished for defection more than male partners while male partners were punished for cooperation more than female partners. This replicates the findings of Experiment 1.

Neither the main effect of participant gender, $F(1,158) = 1.09$, $p = .298$, $\eta_p^2 < .01$, nor the interactions between partner behavior and participant gender, $F(1,158) = 1.51$, $p = .221$, $\eta_p^2 < .01$, between partner gender and participant gender, $F(1,158) = 1.71$, $p = .193$, $\eta_p^2 = .01$, and among partner behavior, partner gender, and participant gender, $F(1,158) = 0.05$, $p = .828$, $\eta_p^2 < .01$, were significant.

As explained above, these results should not be taken at face value because it is necessary to disentangle the direct and indirect effect of facial gender on punishment before it is possible to draw clear conclusions.

Model-based analysis

To analyze the results of Experiment 2, we need four sets of the model trees shown in Figure 1: One set for each combination of partner gender and participant gender. The base model fit the data well, $G^2(4) = 2.62$, $p = .624$.

Parameter C representing the probability that participants cooperate with the partners is shown in the middle panel of Figure 3. Female participants cooperated significantly more with female partners than with male partners, $\Delta G^2(1) = 7.03$, $p = .008$, $w = 0.03$, while male participants cooperated with an equal probability with female and male partners, $\Delta G^2(1) = 0.61$, $p = .436$, $w < 0.01$.

The punishment parameters are shown in the middle panel of Figure 4. As in Experiment 1, moralistic punishment was more common than hypocritical punishment, $\Delta G^2(4) = 682.75$, $p < .001$, $w = 0.33$. Moreover irrational punishment occurred with the smallest probability, and was less common than antisocial punishment, $\Delta G^2(4) = 243.46$, $p < .001$, $w = 0.20$.

As in Experiment 1, partner gender had no effect on the probability of moralistic punishment or any other type of punishment when the punishment parameters were expressed as conditional probabilities. Female participants imposed an equal amount of moralistic punishment on female and male partners, $\Delta G^2(1) = 1.41$, $p = .235$, $w = 0.01$. They also did not discriminate between female and male partners when applying

hypocritical punishment, $\Delta G^2(1) = 0.45$, $p = .500$, $w < 0.01$, irrational punishment, $\Delta G^2(1) = 0.50$, $p = .477$, $w < 0.01$, and antisocial punishment, $\Delta G^2(1) = 0.14$, $p = .710$, $w < 0.01$.

Male participants showed a very similar pattern. They imposed moralistic punishment on female and male partners with an equal probability, $\Delta G^2(1) = 0.61$, $p = .435$, $w < 0.01$.

They also did not discriminate between female and male partners when applying hypocritical punishment, $\Delta G^2(1) = 1.43$, $p = .232$, $w = 0.01$, irrational punishment, $\Delta G^2(1) = 0.17$, $p = .679$, $w < 0.01$, and antisocial punishment, $\Delta G^2(1) = 2.77$, $p = .096$, $w = 0.02$.

Female participants were more likely to impose moralistic punishment, $\Delta G^2(2) = 25.69$, $p < .001$, $w = 0.06$, hypocritical punishment, $\Delta G^2(2) = 11.33$, $p = .003$, $w = 0.04$ and irrational punishment, $\Delta G^2(2) = 21.83$, $p < .001$, $w = 0.06$, on their partners than male participants, while there was no difference in the use of antisocial punishment between female and male participants, $\Delta G^2(2) = 4.50$, $p = .105$, $w = 0.03$.

Harshness of moralistic punishment

The analysis of the harshness of moralistic punishment (Table 2) was again restricted to the subsample of participants who used moralistic punishment at least once. As in Experiment 1, the main effect of partner gender was not significant, $F(1,133) = 1.73$, $p = .191$, $\eta_p^2 = .01$. There was no main effect of participant gender, $F(1,133) = 0.01$, $p = .912$, $\eta_p^2 < .01$, and no interaction between partner gender and participant gender, $F(1,133) = 1.08$, $p = .301$, $\eta_p^2 < .01$.

Discussion

As expected, the results obtained in the female sample were very similar to those obtained in Experiment 1. Female participants tended to cooperate more with female partners, but they did impose an equal amount of moralistic punishment on both female and male partners. By recruiting an equally large male sample, it was possible to examine whether male participants would show a different pattern of results. Indeed, different from female participants, the male participants did not show a clear preference for cooperating with female partners, but instead cooperated equally with female and male partners. Just as the female participants, the male participants did not discriminate between female and male partners when applying moralistic punishment or any other form of punishment.

Interestingly, female participants were more likely to impose moralistic punishment on their partners than male participants. Given that moralistic punishment was costly and did

not provide direct benefits for participants, this aspect of the results is consistent with the assumption that men try to maximize their profits while women are more influenced by the interpersonal aspects of social-dilemma games (Kahn et al., 1971). Female participants also had a higher probability of showing hypocritical and irrational punishment. The only exception to this pattern was antisocial punishment, for which no gender differences were observed.

Experiment 3

An important question that is as yet unanswered is how the participants were affected by being moralistically punished by their partners. In Experiments 1 and 2, the partners punished the participants moralistically for failing to reciprocate cooperation. The moralistic punishment by the partners served to create a realistic experience for the participants because it is known that moralistic punishment is common in interactions with real human partners (Carpenter, 2007; Falk et al., 2005). However, it is possible to speculate that the moralistic punishment by the partners may have influenced the results in several ways.

First, it seems possible that participants may have cooperated with their partners because they feared that they would be punished for defection. This introduces some ambiguity into the interpretation of the effects of facial gender on cooperation observed in Experiments 1 and 2. As yet, it seems unclear whether female participants cooperated more with female partners than with male partners because they had an intrinsic motivation to cooperate with them or because they feared the female partners' moralistic punishment more than the male partners' moralistic punishment.

A similar ambiguity affects the interpretation of the finding of Experiment 2 that female participants are more likely than male participants to moralistically punish their partners. It seems possible to speculate that the moralistic punishment by the partners may have created an ad-hoc norm that led participants to also punish their partners moralistically although they would not have adopted this behavior spontaneously. Therefore, it seems conceivable that female participants did not apply more moralistic punishment than male participants because they were intrinsically motivated to do so, but because they were more easily influenced by their partners' example than men, and more likely to uncritically adopt the ad-hoc norm of moralistic punishment that was implicitly communicated by the partners' behaviors.

To eliminate this ambiguity in the interpretation of the results, it seemed necessary to repeat Experiment 2 but without punishment by partners. Experiment 3 was identical to

Experiment 2 with the only difference being that the punishment option was only available to the participants, and not to the partners. To the extent that the participants' cooperation was motivated by the fear of moralistic punishment, the willingness to cooperate reflected in parameter C can be expected to decrease in Experiment 3 compared to Experiment 2. The main question was whether the effects of gender on cooperation and punishment obtained in Experiment 2 could be replicated under these conditions.

Method

Participants

Eighty-two women and eighty men recruited at Heinrich Heine University participated for monetary compensation or course credit. Their ages ranged between 15 and 49 years ($M = 23$, $SD = 5$). The data set of one female student had to be excluded from data analysis because she had participated twice.

Material, Procedure and Design

Material, Procedure and Design of Experiment 3 were identical to Experiment 2 with the only exception that the punishment option was only available to the participants and not to the partners.

Results

Aggregated amount of money spent to punish the partners

Sixty-three of the 81 female participants (78%) and 50 of the 80 male participants (63%) punished the partners' defection at least once. Moreover, 24 female participants (30%) and 17 male participants (21%) punished the partners' cooperation at least once.

The aggregated amount of money spent to punish the partners is reported in Table 1. Participants spent more money to punish defecting partners than to punish cooperating partners, $F(1,159) = 73.90$, $p < .001$, $\eta_p^2 = .32$. The main effect of partner gender was not significant, $F(1,159) = 3.08$, $p = .081$, $\eta_p^2 = .02$. The interaction between partner behavior and partner gender was not significant, $F(1,159) = 0.09$, $p = .761$, $\eta_p^2 < .01$. Neither the main effect of participant gender, $F(1,159) = 1.19$, $p = .276$, $\eta_p^2 < .01$, nor the interactions between partner behavior and participant gender, $F(1,159) = 0.14$, $p = .708$, $\eta_p^2 < .01$, between partner gender and participant gender, $F(1,159) = 1.32$, $p = .252$, $\eta_p^2 < .01$, and among partner gender, partner behavior, and participant gender, $F(1,159) = 1.68$, $p = .196$,

$\eta_p^2 = .01$, were significant. Again, these findings should only be cautiously interpreted because they provide only coarse information of the effects of gender on punishment. A more fine-grained analysis is needed before drawing further conclusions.

Model-based analysis

The same base model was used as in Experiment 2, which fit the data well, $G^2(4) = 1.71$, $p = .788$.

Parameter *C* representing the probability of the participants' cooperation is shown in the right panel of Figure 3. Cooperation was noticeably reduced in Experiment 3 in comparison to Experiments 1 and 2, but cooperation was clearly above zero. As in Experiment 2, female participants cooperated more with female partners than with male partners, $\Delta G^2(1) = 15.50$, $p < .001$, $w = 0.05$, while male participants were equally likely to cooperate with female and male partners, $\Delta G^2(1) = 0.07$, $p = .796$, $w < 0.01$.

The punishment parameters are shown in the lower panel of Figure 4. As in Experiments 1 and 2, moralistic punishment was more common than hypocritical punishment, $\Delta G^2(4) = 653.32$, $p < .001$, $w = 0.32$. In contrast to Experiments 1 and 2, antisocial punishment was as rare as irrational punishment, $\Delta G^2(4) = 0.82$, $p = .936$, $w = 0.01$.

As in Experiments 1 and 2, partner gender had no effect on the probability of moralistic punishment or any other type of punishment when the punishment parameters were expressed as conditional probabilities. Female participants imposed an equal amount of moralistic punishment on female and male partners, $\Delta G^2(1) = 0.45$, $p = .503$, $w < 0.01$. They also did not discriminate between female and male partners when applying hypocritical punishment, $\Delta G^2(1) = 0.66$, $p = .415$, $w = 0.01$, irrational punishment, $\Delta G^2(1) = 0.18$, $p = .670$, $w < 0.01$, and antisocial punishment, $\Delta G^2(1) = 0.05$, $p = .820$, $w < 0.01$. Male participants showed a similar pattern. They imposed moralistic punishment with an equal probability on female and male partners, $\Delta G^2(1) = 1.07$, $p = .301$, $w = 0.01$. They also did not discriminate between female and male partners when applying hypocritical punishment, $\Delta G^2(1) = 0.04$, $p = .848$, $w < 0.01$, irrational punishment, $\Delta G^2(1) = 0.02$, $p = .893$, $w < 0.01$, and antisocial punishment, $\Delta G^2(1) = 0.01$, $p = .913$, $w < 0.01$.

Female participants were more likely to impose moralistic punishment, $\Delta G^2(2) = 20.87$, $p < .001$, $w = 0.06$, hypocritical punishment, $\Delta G^2(2) = 37.89$, $p < .001$, $w = 0.08$, irrational

punishment, $\Delta G^2(2) = 34.54$, $p < .001$, $w = 0.07$, and antisocial punishment, $\Delta G^2(2) = 51.24$, $p < .001$, $w = 0.09$, on their partners than male participants.

Harshness of moralistic punishment

As in Experiment 2, the main effect of participant gender missed the conventional level of statistical significance, $F(1,91) = 3.59$, $p = .061$, $\eta_p^2 = .04$. There was neither a main effect of partner gender, $F(1,91) = 0.78$, $p = .379$, $\eta_p^2 < .01$, nor an interaction between participant gender and partner gender, $F(1,91) = 1.15$, $p = .286$, $\eta_p^2 = .01$.

Discussion

The first important finding of Experiment 3 is that, overall, the pattern of results was very similar to that obtained in Experiment 2. First, female participants cooperated more with female partners than with male partners while male participants did not discriminate between female and male partners. If anything, the female participants' preference for cooperating more with female partners than with male partners was more pronounced in Experiment 3 in comparison to Experiment 2, which suggests that fear of punishment does not play a role in explaining this preference. Second, participants again applied moralistic punishment with a higher probability than any other type of punishment. This disconfirms the idea that the participants' moralistic punishment in Experiments 1 and 2 was merely the result of them following the example of the partners. Even when the participants did not experience moralistic punishment by the partners, they still used the punishment option primarily to apply moralistic punishment. This suggests that participants do not simply mimic the behavior of their partners because the high propensity for punishing moralistically is independent of whether or not participants can observe partners showing this behavior. Third, partner gender had no direct effects on punishment, as in Experiments 1 and 2. The participants imposed moralistic punishment (and any other type of punishment) with an equal probability on female and male partners. Fourth, we replicated the finding that female participants were more likely to impose moralistic punishment (and any other form of costly punishment) on their partners than male participants.

However, there are also important differences to Experiment 2. First, the participants' willingness to cooperate decreased in Experiment 3 in comparison to Experiment 2. This suggests that the moralistic punishment of the partners served to enhance cooperation in Experiments 1 and 2 and thus confirms the view that moralistic punishment is indeed

effective in increasing cooperation. Thus, the cooperation of both female and male participants in Experiments 1 and 2 seems to have been partly motivated by the fear of being moralistically punished. However, cooperation was not completely abolished, which suggests that cooperation was not solely motivated by the fear of punishment. In Experiment 3, participants could increase their payoff by refusing cooperation, and there was no punishment for defection (and, therefore, no monetary incentive for cooperation). The rational choice in terms of maximizing one's own payoff would be to always defect. It seems noticeable that a comparatively high level of cooperation was sustained even in the absence of moralistic punishment by the partners.

Although it is difficult to compare the results of two separate experiments directly, it seems noteworthy that the decrease in cooperation in Experiment 3, in comparison to Experiment 2, is somewhat more pronounced for male in comparison to female participants. This difference between experiments suggests that the cooperation of male participants is more strongly motivated by monetary incentives (and, therefore, more influenced by moralistic punishment) while female participants have a higher intrinsic motivation to cooperate with their partners.

There are also differences in punishment between Experiment 3 and Experiment 2. Most noticeably, the probability of antisocial punishment was much higher than the probability of irrational punishment in both Experiments 1 and 2, but antisocial punishment markedly decreased, and was not significantly different from irrational punishment in Experiment 3. This is in line with the idea that antisocial punishment is, in part, motivated by the retaliation of anticipated moralistic punishment of the partners (Nikiforakis, 2008). Uncooperative participants may have anticipated the moralistic punishment of cooperative partners in Experiments 1 and 2, which may have motivated them to prospectively retaliate the punishment. In Experiment 3, in contrast, there is no reason to anticipate moralistic punishment by the partners, and, therefore, less reason for antisocial punishment.

General Discussion

Common gender stereotypes imply that women are kind, submissive, gentle, and emotional, while men are perceived as being competitive, dominant, aggressive, and rational (e.g., Prentice & Carranza, 2002; Rosenkrantz, Vogel, Bee, Broverman, & Broverman, 1968; Williams & Bennett, 1975). In line with these stereotypes people expect women to be more cooperative and less competitive than men in social interactions (Balliet et al., 2011; Glick & Fiske, 2001; Orbell, Dawes, & Schwartz-Shea, 1994). In the present

study, we wanted to test the hypothesis that people tend to punish stereotype-incongruent behaviors in a social-dilemma game with a punishment option. Consistent with this hypothesis, we observed in Experiment 1 that, aggregated over trials, participants spent more money to punish the defection of female partners than to punish the defection of male partners while they spent more money to punish the cooperation of male partners than to punish the cooperation of female partners. At first glance, this finding could be taken as evidence that participants punish stereotype-violating behaviors more than stereotype-confirming behaviors.

However, a more fine-grained analysis showed that this effect of gender on punishment was indirect rather than direct, and attributable to the fact that participants cooperated more with female partners, which provided them with more opportunity for moralistic punishment. When the influence of this gender-specific increase in cooperation was taken into account by expressing the different types of punishment as conditional probabilities (i.e., conditional on whether cooperation occurred), no direct effect of partner gender on any form of punishment could be observed. It thus seems possible to conclude that women may indeed experience a higher total amount of punishment for uncooperative behaviors than men, but only because they are met with more cooperation. When cooperation is not reciprocated, people tend to punish this lack of reciprocation with moralistic punishment. However, they do not punish female partners more for failing to reciprocate their cooperation than male partners. The higher amount of punishment directed at the female partners is completely due to the higher rate of cooperation directed at the female partners.

Experiment 2 showed that the increased cooperation towards female partners is primarily due to the fact that female participants tended to cooperate more with female partners while male participants did not discriminate between female and male partners. Experiment 3 clarified that this increase in cooperation is not due to an increased fear of moralistic punishment by female partners because women's preference for cooperating with female partners was clearly observed (and descriptively even more pronounced) in the absence of moralistic punishment by the partners.

By examining both female and male participants, it was possible to test the assumption that women are indeed less interested in maximizing profit and more socially oriented than men (e.g., Andreoni & Vesterlund, 2001; Eckel & Grossman, 1998, 2008). At first glance, we may expect that female participants should show more cooperation than male participants if this hypothesis is true. However, participants were moralistically punished for

defection in Experiment 2. Therefore, some participants may have chosen to cooperate in an attempt to increase their payoffs by avoiding moralistic punishment. Experiment 3 provides a clearer picture on this issue. Participants did not receive any punishment in this experiment so that there was no need to cooperate out of the fear of being punished. Given that defection always led to a higher payoff than cooperation in the prisoner's dilemma game, the only rational choice (in terms of maximizing one's own payoff) in Experiment 3 was to defect. In this situation, female participants did indeed cooperate more with their partners than male participants although this gender difference was confined to interactions with female partners. This finding provides some evidence for the assumption that female participants were somewhat less interested in maximizing their own payoffs, and were somewhat more interested in the social aspects of the task. This interpretation is further strengthened by the finding that female participants were more likely than male participants to impose punishment on the partners. Given that participants interacted with each partner only once, and punishment was costly, the rational choice (in terms of maximizing one's payoff) would have been not to use the punishment option at all because punishing the partners had no direct benefit for the participants. In contrast to this prediction of rational choice theory, both female and male participants used the punishment option with a high probability, primarily to punish the partners moralistically for failing to reciprocate cooperation. However, female participants were somewhat more likely to punish their partners (moralistically or otherwise) than male participants in both Experiments 2 and 3. Thus, there is indeed some evidence in the present results that women are more socially oriented than men, consistent with previous findings (Andreoni & Vesterlund, 2001; Eckel & Grossman, 1998, 2008). Experiment 3 showed that female participants did not only follow the example of their partners, but were intrinsically motivated to punish their partners, because the gender differences in punishment were reliably obtained even in the absence of moralistic punishment by the partners.

Although participants did not know that they would not interact with real human partners beforehand a potential concern is that participants played the social-dilemma game with preprogrammed partners rather than real human partners. However, this gave us the opportunity to experimentally control the partners' behaviors. Furthermore, the participants' behaviors were incentivized by real monetary consequences. It seems possible to speculate that participants did not respond in the same way as when interacting with real human partners, but this idea is seriously challenged by the findings that (a) participants showed a high level of cooperation even in Experiment 3 in which there was no monetary

incentive to cooperate, but a clear monetary incentive to defect, and that (b) participants used the punishment option with a high probability in all three experiments although punishing was costly and provided no direct benefit for the participants. These behaviors—that can be considered “irrational” from an economic perspective—suggest that participants responded in a very similar way to their partners as in games with real human partners (Carpenter, 2007; Falk et al., 2005).

To summarize, we found in Experiment 1 that participants spent more money for the punishment of defection when they interacted with female partners than when they interacted with male partners while the opposite was true for the punishment of cooperation. Although this finding is, at first glance, consistent with the idea that the violation of gender stereotypes receives special punishment, this effect was entirely due to an indirect rather than to a direct effect of facial gender on punishment. Female partners were met with more cooperation than male partners, and participants generally tended to punish their partners for failing to reciprocate their own cooperation. Experiments 2 and 3 clarified that female, but not male, participants tended to cooperate more with female partners than with male partners, regardless of whether or not they had to fear their partners’ moralistic punishment. Furthermore, women tended to punish their partners with a higher probability than men.

Figures

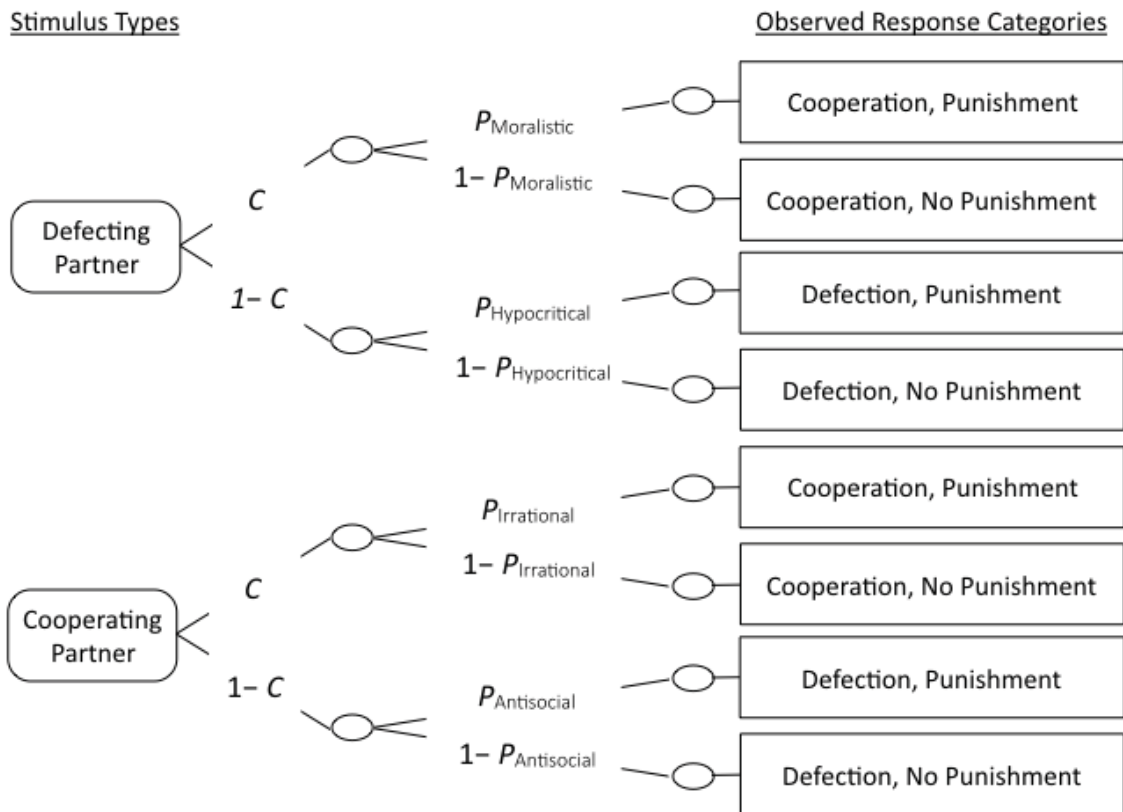


Figure 1. Multinomial model of cooperation and punishment. Rounded rectangles on the left side represent the partners who either defect or cooperate. Rectangles on the right side show the observable response categories. Participants cooperate with probability C . The P -parameters represent the conditional probabilities of four types of punishment (moralistic, hypocritical, irrational, and antisocial punishment).

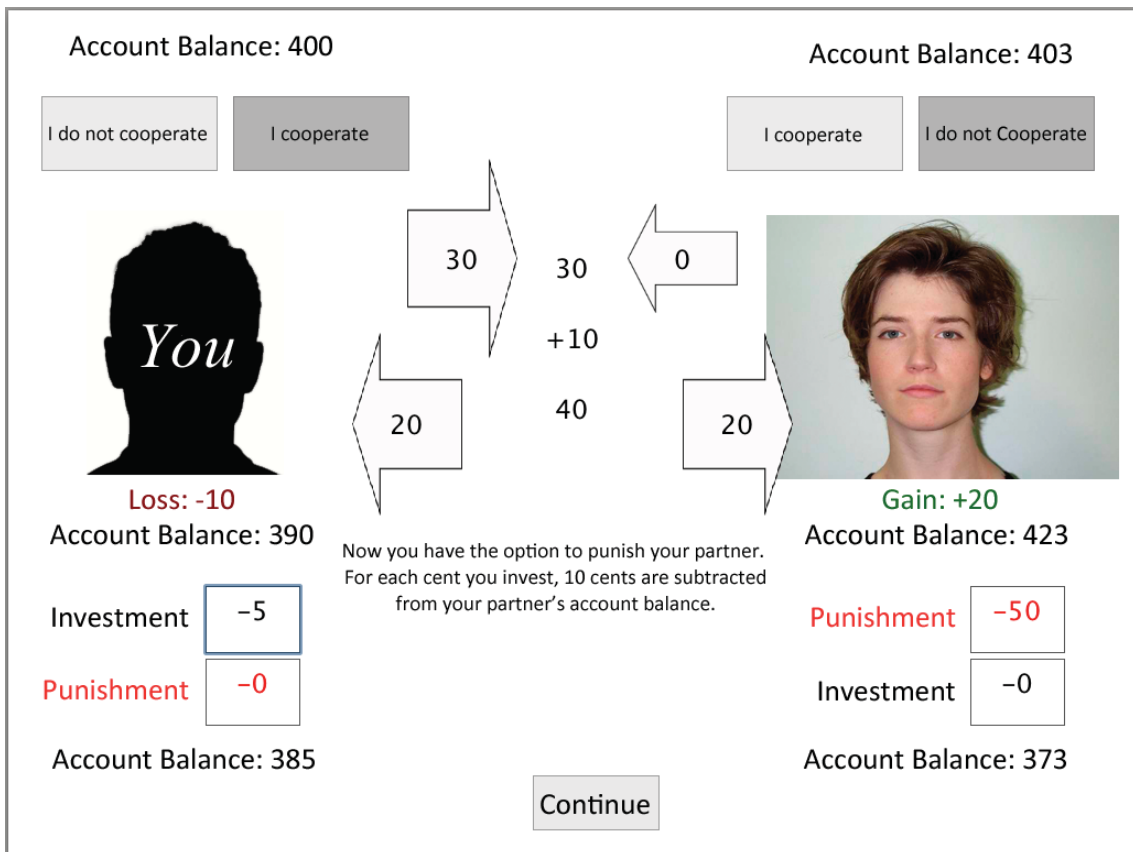


Figure 2. Screenshot of the prisoner's dilemma game that was used in Experiments 1 and 2. The only difference in Experiment 3 was that the punishment option was only available to the participants and not to the partners. The facial photograph of the partner was taken from the Center for Vital Longevity Face Database (Minear & Park, 2004).

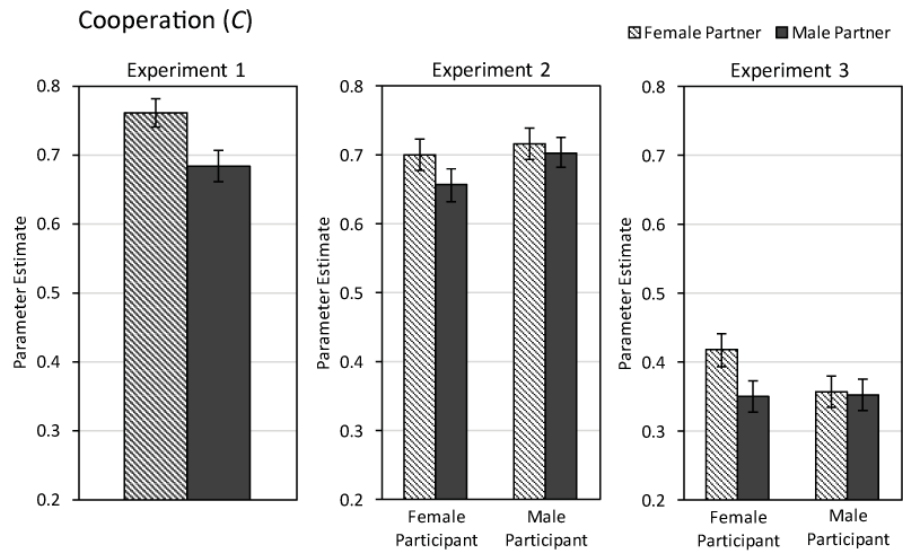


Figure 3. Parameter estimates for parameter *C*, representing the probability that a participant decides to cooperate rather than to defect, as a function of the partner's gender in Experiments 1, 2, and 3. In Experiment 1, the parameters represent the estimates for female and male participants combined. In Experiments 2 and 3, cooperation was analyzed separately for female and male participants. The error bars represent the .95-confidence intervals.

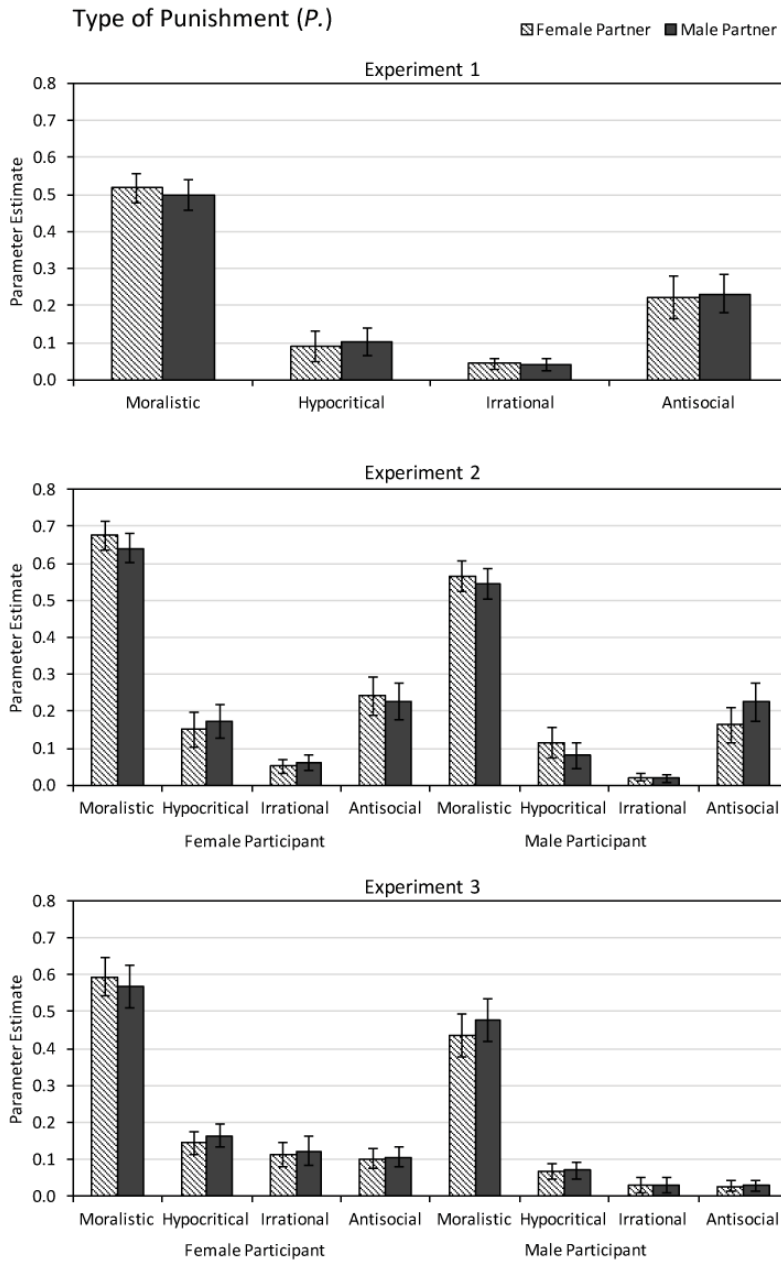


Figure 4. Parameter estimates for punishment parameters P . representing the conditional probabilities that a participant decides to punish a partner as a function of type of punishment (moralistic, hypocritical, irrational and antisocial), and the partner’s gender (in Experiments 1, 2, and 3). Note that the type of punishment depends on the participants’ decision to defect or to cooperate in the first stage of the game. In Experiment 1, the parameters represent the estimates for female and male participants combined because there were too few male participants to analyze their responses separately. In Experiments 2 and 3, punishment was analyzed separately for female and male participants. The error bars represent the .95-confidence intervals.

Tables

Table 1: The aggregated amount of money spent to punish the partners in cent (summarized across trials and averaged across participants) as a function of partner behavior (defection or cooperation) and partner gender (female or male). In Experiments 2 and 3, results are presented separately for female and male participants.

| | Defecting Partner | | | | Cooperating Partner | | | |
|---------------------|-------------------|---------------|--------------|---------------|---------------------|---------------|--------------|---------------|
| | Female Partner | | Male Partner | | Female Partner | | Male Partner | |
| | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) |
| Experiment 1 | 15.60 | (1.68) | 13.45 | (1.48) | 2.60 | (0.65) | 3.95 | (0.94) |
| Experiment 2 | | | | | | | | |
| Female Participants | 20.84 | (1.99) | 19.36 | (1.88) | 4.00 | (1.15) | 4.11 | (1.11) |
| Male Participants | 17.23 | (1.70) | 16.76 | (1.74) | 2.75 | (0.81) | 4.23 | (1.22) |
| Experiment 3 | | | | | | | | |
| Female Participants | 8.26 | (0.96) | 8.15 | (1.05) | 1.95 | (0.59) | 2.42 | (0.69) |
| Male Participants | 6.73 | (1.28) | 8.05 | (1.48) | 0.63 | (0.22) | 1.01 | (0.37) |

Table 2: The harshness of moralistic punishment in terms of the amount of money spent to punish the partners if participants decided to punish them moralistically.

| | Defecting Partner | | | |
|---------------------|-------------------|---------------|--------------|---------------|
| | Female Partner | | Male Partner | |
| | <i>M</i> | (<i>SE</i>) | <i>M</i> | (<i>SE</i>) |
| Experiment 1 | 3.67 | (0.23) | 3.59 | (0.22) |
| Experiment 2 | | | | |
| Female Participants | 3.69 | (0.21) | 3.95 | (0.23) |
| Male Participants | 3.84 | (0.20) | 3.87 | (0.24) |
| Experiment 3 | | | | |
| Female Participants | 2.65 | (0.22) | 2.91 | (0.26) |
| Male Participants | 3.56 | (0.38) | 3.53 | (0.35) |

References

- Andreoni, J., & Vesterlund, L. (2001). Which is the Fair Sex? Gender Differences in Altruism. *The Quarterly Journal of Economics*, *116*, 293–312.
- Balliet, D., Li, N., Macfarlan, S., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin*, *137*, 881–909.
- Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 197–215.
- Bell, R., Buchner, A., Kroneisen, M., & Giang, T. (2012). On the flexibility of social source memory: A test of the emotional incongruity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1512–1529.
- Bell, R., Buchner, A., & Musch, J. (2010). Enhanced old-new recognition and source memory for faces of cooperators and defectors in a social-dilemma game. *Cognition*, *117*, 261–275.
- Bröder, A., & Meiser, T. (2007). Measuring source memory. *Zeitschrift für Psychologie / Journal of Psychology*, *215*, 52–60.
- Carpenter, J. (2007). The demand for punishment. *Journal of Economic Behavior & Organization*, *62*, 522–542.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum, Inc.
- Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality & Social Psychology Bulletin*, *15*, 543–558.
- Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *The Economic Journal*, *108*, 726–735.
- Eckel, C. C., & Grossman, P. J. (2008). Differences in the economic decisions of men and women: Experimental evidence. In C. Plott, & V. Smith (Eds.), *Handbook of Experimental Economics Results* (pp. 509–519). New York: Elsevier.
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models. *Zeitschrift für Psychologie / Journal of Psychology*, *217*, 108–124.
- Falk, A., Fehr, E., & Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, *73*, 2017–2030.

- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.
- Giang, T., Bell, R., & Buchner, A. (2012). Does facial resemblance enhance cooperation? *PLoS ONE, 7*, e47809.
- Glick, P., & Fiske, S. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *The American Psychologist, 56*, 109–118.
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., . . . Ziker, J. (2006). Costly punishment across human societies. *Science, 312*, 1767–1770.
- Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. *Science, 319*, 1362–1367.
- Kahn, A., Hottes, J., & Davis, W. L. (1971). Cooperation and optimal responding in the Prisoner's Dilemma game: Effects of sex and physical attractiveness. *Journal of Personality and Social Psychology, 17*, 267–279.
- Klauer, K., Stahl, C., & Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 680–703.
- Mieth, L., Bell, R., & Buchner, A. (2016). Facial likability and smiling enhance cooperation, but have no direct effect on moralistic punishment. *Manuscript submitted for publication*.
- Miner, M., & Park, D. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*, 630–633.
- Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*, 42–54.
- Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics, 92*, 91–112.
- Orbell, J., Dawes, R., & Schwartz-Shea, P. (1994). Trust, social categories, and individuals: The case of gender. *Motivation and Emotion, 18*, 109–128.
- Prentice, D., & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly, 26*, 269–281.
- Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95*, 318–339.

- Rosenkrantz, P., Vogel, S., Bee, H., Broverman, I., & Broverman, D. M. (1968). Sex-role stereotypes and self-concepts in college students. *Journal of Consulting and Clinical psychology, 32*, 287–295.
- Solnick, S., & Schweitzer, M. (1999). The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Processes, 79*, 199–215.
- Todorov, A. (2008). Evaluating faces on trustworthiness. *Annals of the New York Academy of Sciences, 1124*, 208–224.
- Williams, J. E., & Bennett, S. M. (1975). The definition of sex stereotypes via the adjective check list. *Sex Roles, 1*, 327–337.
- Wilson, R. K., & Eckel, C. C. (2006). Judging a book by its cover: Beauty and expectations in the trust game. *Political Research Quarterly, 59*, 189–202.

Erklärung an Eides Statt

Hiermit versichere ich an Eides Statt, dass ich die Arbeit mit dem Titel „Der Einfluss von Erwartungen an soziale Interaktionspartner auf Kooperations- und Bestrafungsverhalten“ selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ verfasst habe. Ich versichere insbesondere:

- (1) Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt.
- (2) Alle wörtlich oder dem Sinn nach aus anderen Texten entnommenen Stellen habe ich als solche kenntlich gemacht; dies gilt für gedruckte Texte ebenso wie für elektronische Ressourcen.
- (3) Die Arbeit habe ich in der vorliegenden oder einer modifizierten Form noch nicht als Dissertation vorgelegt – sei es an der Heinrich-Heine-Universität Düsseldorf oder an einer anderen Universität.

Datum: 20. Juli 2016

Name: Laura Mieth

Unterschrift: