Diplomarbeit am Fachbereich Psychologie
Abteilung Allgemeine Psychologie II

Neuronale und behaviorale Erklärungsansätze für das Phänomen der altruistischen Bestrafung

Jörg Groß und Franziska Schuricht
Goethe-Universität, Frankfurt am Main

Erstgutachter: Prof. Dr. Alexander Strobel
Zweitgutachter: Prof. Dr. Sabine Windmann

Franziska Schuricht
Matrikelnr: 300 8122
Dortelweilerstraße 6
60389 Frankfurt

Jörg Groß
Matrikelnr: 300 7521
Throner Straße 15
60385 Frankfurt

Zusammenfassung

Viele Menschen sind bereit eigene Kosten in Kauf zu nehmen, um andere für normverletzendes Verhalten zu bestrafen. Dieses Phänomen wird altruistische Bestrafung genannt und als Erklärung für die Entstehung und Erhaltung von sozialen Normen in menschlichen Gesellschaften herangezogen. Ausgehend von einer Fairnessnorm, die intersubjektiv die Bewertung von sozialen Interaktionen ermöglicht, vermuten wir, dass Menschen dazu bereit sind, auch unter Kostenaufwand und unabhängig davon, ob sie selbst von einem Normbruch betroffen sind (First Person) oder diesen lediglich beobachten (Third Party), bei unfairem Verhalten zu intervenieren. In der vorliegenden Untersuchung sollte diese Annahme auf behavioraler Ebene geprüft sowie die zu Grunde liegenden neuronalen Mechanismen unter Einsatz von funktioneller Magnetresonanztomographie untersucht werden. Wir ließen 24 Studierende der Universität Frankfurt ein Dictator Game sowohl aus einer First Person- als auch einer Third Party Perspektive spielen. Je nach Durchgang hatten sie die Möglichkeit Aufteilungen eines Geldbetrages unter Kosteneinsatz zu bestrafen oder deren Fairness zu beurteilen. Auf behavioraler Ebene stellten wir fest, dass sich das Ausmaß an altruistischer Bestrafung kaum zwischen den Perspektiven unterschied und eng an die abgegebenen Fairnessurteile geknüpft war. Auf neuronaler Ebene zeigte sich eine erhöhte Aktivität des dorsolateralen präfrontalen Cortex (DLPFC), des anterioren cingulären Cortex (ACC), des Precuneus und der Insula, wenn die Probanden bestraften. Wir vermuten, dass der DLPFC und der ACC kognitive Kontrollfunktionen bei der altruistischen Handlung übernehmen, während die Insula bei emotionalen Prozessen, wie Ärger, eine Rolle spielt. Weiterhin fanden wir, dass der Precuneus neben ich-bezogenen Handlungen bei direkter Betroffenheit auch mit dem aktiven Eingriff in eine beobachtete soziale Interaktion assoziiert ist. Der Nucleus caudatus war bei Bestrafung und Fairnessurteil vergleichbar aktiv. Dieses Ergebnis spricht gegen eine reine Belohnungsfunktion dieser Hirnregion bei altruistischer Bestrafung, wie von anderen Autoren vermutet. In unserer Studie konnten wir weiterhin einen positiven Zusammenhang zwischen dem Persönlichkeitsmerkmal Empathiefähigkeit und der Aktivierung der rechten Insula finden, wenn aus der Third Party Perspektive bestraft wurde. Im Rahmen einer evolutionspsychologischen Theorie wird davon ausgegangen, dass Empathie die Fähigkeit bietet Fairness losgelöst der eigenen Interessen zu bewerten. Dies wird als mögliche Grundlage für die Entstehung einer Fairnessnorm angeführt, welche wiederum durch altruistische Bestrafung aufrecht erhalten werden kann.

Inhaltsverzeichnis

1 Einleitung **5**

1.1 Altruismus 7

1.1.1 Biologischer versus psychologischer Altruismus 7

1.1.2 Altruismus in der Evolutionspsychologie 8

1.1.3 Ultimate Erklärungsansätze für altruistisches Verhalten 9

1.1.3.1 Altruistisches Verhalten unter verwandten Individuen 9

1.1.3.2 Altruistisches Verhalten unter nicht verwandten Individuen . . 10

1.2 Altruistische Bestrafung 16

1.2.1 Nutzen und Kosten von altruistischer Bestrafung 17

1.2.2 Proximate Erklärungsansätze für altruistische Bestrafung 19

1.2.2.1 Soziale Normen 19

1.2.2.2 Emotionen 20

1.2.2.3 Empathie 22

1.2.2.4 Ungerechtigkeitssensibilität 23

1.3 Altruistische Bestrafung und Ökonomie 24

1.3.1 Eigennutz in der Ökonomie – Der Mensch als Homo oeconomicus 24

1.3.2 Fairness in der Ökonomie – Die Theorie der sozialen Präferenzen 25

1.3.2.1 Ungleichheitsaversion 25

1.3.2.2 Reziproke Fairness 26

1.3.2.3 Die Theorie der sozialen Präferenzen – Zusammenfassung . . . 26

1.3.3 Das Prinzip der Spieltheorie 27

1.3.4 Paradigmen der Spieltheorie 27

1.3.4.1 Das Dictator Game 28

1.3.4.2 Das Dictator Game mit Bestrafungsoption 28

1.4 Altruistische Bestrafung, Ökonomie und Neuroforschung 29

1.4.1 Das neue Forschungsfeld der Neuroökonomie 29

1.4.2 Multiple Ansätze zur Entscheidungsfindung 30

1.4.3 Die Methode der funktionellen Magnetresonanztomographie 32

1.4.4 Neuronale Korrelate von altruistischer Bestrafung 32
1.4.4.1 Die Insula 33
1.4.4.2 Das Striatum 34
1.4.4.3 Der frontale Neocortex 35
1.4.4.4 Der anteriore cinguläre Cortex 36
1.5 Altruistische Bestrafung in der vorliegenden Studie 37
1.5.1 Fragestellung und Hypothesen 38
1.5.1.1 Verhaltensebene 38
1.5.1.2 Neuronale Ebene 38
1.5.1.3 Differentialpsychologische Aspekte 39

2 Methode **40**
2.1 Stichprobe 40
2.2 Versuchsablauf und Versuchsmaterialien 41
2.2.1 Fragebogenbatterie 41
2.2.2 Das Dictator Game mit Bestrafungsoption und Fairnessurteil 42
2.3 Technische Umsetzung 44
2.3.1 Visuelle Stimulation und Response-Box 44
2.3.2 Imaging Parameter 46
2.4 Statistische Analyse 46
2.4.1 Verhaltensebene 46
2.4.2 Neuronale Ebene 47
2.4.2.1 Preprocessing 47
2.4.2.2 First-Level Analyse 48
2.4.2.3 Second-Level Analyse 48
2.4.2.4 Event-Related Averages 50
2.4.3 Differentialpsychologische Analyse 50

3 Ergebnisse **51**
3.1 Verhaltensebene 51
3.1.1 Verhaltenskonsistenz 53
3.1.2 Reaktionszeiten 55

3.2 Neuronale Ebene 56
3.2.1 Second-Level Analyse 57
3.2.1.1 Einzelkontrast: Bestrafung versus keine Bestrafung 57
3.2.1.2 Kontrast: First Person versus Third Party 57
3.2.1.3 Kontrast: Bestrafungsdurchgänge versus Fairnessdurchgänge . 59
3.2.2 Ereignisbezogene Aktivitätsunterschiede 62
3.2.2.1 Insula 62
3.2.2.2 Dorsolateraler präfrontaler Cortex 64
3.2.2.3 Precuneus 65
3.2.2.4 Dorsales Striatum 67
3.2.2.5 Anteriorer cingulärer Cortex 68
3.3 Differentialpsychologische Ergebnisse 69
3.3.1 Ungerechtigkeitssensibilität 69
3.3.2 Empathie 70
3.3.3 Reziprozität 71

4 Diskussion **71**
4.1 Ergebniszusammenfassung 71
4.2 Einbettung der Ergebnisse in die bisherige Befundlage 73
4.3 Ein Entstehungsmodell der altruistischen Bestrafung – Resümee 77
4.4 Grenzen der Studie und Ausblick 79

Literatur **83**

1 Einleitung

Menschen arbeiten ehrenamtlich in sozialen Einrichtungen, spenden zum Teil beachtliche Summen für wohltätige Zwecke und bei Naturkatastrophen oder anderen Notfallsituationen riskieren sie mitunter sogar das eigene Leben für die Rettung des Lebens eines anderen Menschen. Auch in ganz alltäglichen Situationen begegnen uns Menschen, welche spontan anderen helfen oder ihnen etwas Gutes tun. Personen in Eile finden dennoch Zeit einem nach dem Weg fragenden Fremden Auskunft zu geben, Schüler und Studierende überlassen selbst ausgearbeitete Skripte Freunden und Kommilitonen, eine blinde Person wird über die stark befahrene Straße geführt, einer kleinen Person die Ware aus dem obersten Regal gereicht, für einen verhinderten Kollegen wird kurzfristig eingesprungen.

Menschen sind gewillt, selbst knappe und wertvolle Ressourcen mit anderen Menschen zu teilen. Sie verhalten sich kooperativ und hilfsbereit auch ohne den äußeren Druck durch Gesetzesvorgaben (Fehr & Fischbacher, 2004a). Die Gesamtheit dieser informellen, sozialen Regeln zu Gerechtigkeit und Fairness bildet die sozialen Normen einer Gesellschaft. Diese Normen werden in der Regel von den Mitgliedern einer sozialen Gemeinschaft anerkannt und eingehalten. Sie bieten bedeutsame Richtlinien für Verhalten und erleichtern Handlungsentscheidungen sowie die Kommunikation im sozialen Kontext (Elster, 1989). Doch auch das Wissen darum, wegen eines normverletzenden Verhaltens bestraft zu werden, trägt zur Aufrechterhaltung von sozialen Normen bei (Fehr & Gächter, 2002). Diese Bestrafung nützt somit allen kooperierenden Mitgliedern der sozialen Gemeinschaft. Jedoch setzt sie den Einsatz von Kosten der bestrafenden Person voraus; es wird daher von altruistischer Bestrafung gesprochen. Die Bestrafungskosten des Einzelnen sind hierbei meist geringer relativ zum erbrachten Nutzen für die Gruppe. Dadurch lässt sich das evolutionäre Bestehen von altruistischer Bestrafung durch einen Fitnessvorteil in unmittelbarem Bezug auf die ganze Gruppe erklären (Gintis, 2000). Innerhalb der Gruppe aber besitzen kooperierende und bestrafende Mitglieder einen relativen Kostennachteil sowohl gegenüber von nicht kooperierenden *Free Rider* (Fehr & Gächter, 2002, Seite 137) als auch gegenüber von kooperierenden und nicht bestrafenden Mitgliedern. Auf Individualebene scheinen ultimate Modelle den Einsatz von altruistischer Bestrafung somit nicht begründen zu können. Proximate Verhaltensmodelle ziehen daher psychologische Faktoren als Erklärung für den Einsatz von altruistischer Bestrafung in Betracht. Ist eine Person selbst von einem unfairen Verhalten be-

troffen (First Person Perspektive), kann die Genugtuung darüber, sich hierfür rächen zu können, ein Motiv für altruistische Bestrafung bilden. Doch Menschen bestrafen auch dann altruistisch, wenn sie ein ungerechtes Verhalten lediglich aus der Perspektive einer dritten Person (Third Party Perspektive) beobachten (Fehr & Fischbacher, 2004b).

Ziel der vorliegenden Untersuchung ist es, das Phänomen der altruistischen Bestrafung sowohl aus der First Person als auch aus der Third Party Perspektive zu untersuchen und besser verstehen zu können. Die Fairnessnorm der objektiven Gleichverteilung von Ressourcen, wie durch die *Equity Theorie* postuliert (Adams, 1965), bildet dabei den Ausgangspunkt zur Erklärung von altruistischer Bestrafung unabhängig von der Wahrnehmungsperspektive. Die Annahme eines kausalen Zusammenhangs zwischen der Wahrnehmung eines sozialen Normbruchs und dem Einsatz von altruistischer Bestrafung (z.B. Fehr & Gächter, 2002; Fehr, Fischbacher & Gächter, 2003; Fehr & Fischbacher, 2004a; de Quervain, Fischbacher, Treyer, Schellhammer, Buck & Fehr, 2004) soll sowohl auf behavioraler als auch auf neuronaler Ebene empirisch geprüft werden. Weiterhin interessiert uns der mögliche Einfluss von interindividuell variierenden Persönlichkeitsmerkmalen auf das Ausmaß von altruistischer Bestrafung.

Vor der Darstellung von Methode, Ergebnissen und Diskussion der eigenen Studie wird in den folgenden Abschnitten zunächst der theoretische Bezugsrahmen hierfür im Detail aufgeführt. Dabei soll vor allem verdeutlicht werden, was die Antwort auf die Frage, warum altruistisch bestraft wird, so bedeutsam und schwierig zugleich macht. Das Investieren eigener Kosten zu Gunsten anderer scheint weder mit den etablierten Grundsätzen der Evolutionspsychologie noch mit traditionellen Verhaltensmodellen der Ökonomie vereinbar. Hiernach gelten Prinzipien der eigenen Nutzenmaximierung. Es wird daher ein Überblick solcher Erklärungsmodelle gegeben, mit welchen bereits versucht wurde, diese Prinzipien in Einklang mit Altruismus im Allgemeinen sowie altruistischer Bestrafung im Speziellen zu bringen. Die Paradigmen der Spieltheorie bieten dabei eine hervorragende Möglichkeit, solche Situationen experimentell umzusetzen, in denen altruistisch bestraft wird. Die neuronalen Korrelate von gezeigtem Verhalten können wiederum mit Hilfe von modernen bildgebenden Methoden der Neurowissenschaft untersucht werden und den Erklärungsrahmen für Verhalten somit stark erweitern. Basierend auf den Ergebnissen aktueller neuroökonomischer Studien wird eine Auswahl von solchen Regionen des Gehirns näher betrachtet, welche eine Rolle bei altruistischer Bestrafung zu spielen scheinen. Abschließend formulieren wir die Hypothesen und Fragestellungen der eigenen Studie noch einmal ausführlich

und mit Bezug auf den theoretischen Hintergrund. Begonnen wird mit der begrifflichen Spezifikation von Altruismus, unser zu untersuchendes Konstrukts im weiteren Sinne.

1.1 Altruismus

Der Begriff Altruismus leitet sich vom lateinischen *alter*, der Andere, ab und enthält somit bereits einen wichtigen Aspekt dessen Bedeutung. Denn altruistische Verhaltensweisen werden der Gruppe der prosozialen Verhaltensweisen zugeordnet und diese beschreiben Handlungen, welche einer anderen Person einen Nutzen erbringen. Helfendes Verhalten und Kooperation bilden zwei Beispiele für derartige Handlungen und werden in den Ausführungen der vorliegenden Arbeit gegebenenfalls synonym verwendet (Penner, Dovidio, Piliavin & Schroeder, 2004). Beinhaltet die prosoziale Handlung neben dem (materiellen) Nutzen für den Empfänger auch Kosten für den Akteur, sind die definitorischen Kriterien für eine altruistische Handlung erfüllt, nämlich „(...) as being costly acts that confer economic benefits on other individuals“ (Fehr & Fischbacher, 2003, Seite 785).

1.1.1 Biologischer versus psychologischer Altruismus

Handelt ein Mensch altruistisch, so lassen sich hierbei unter anderem zwei verschiedene Aspekte differenzieren: die Intention der Handlung sowie die Handlung an sich. Das Konzept des biologischen Altruismus beschränkt sich auf letzteres. Sind die bereits oben zitierten Merkmale für altruistisches Verhalten erfüllt, bedarf es hierfür keines weiteren definitorischen Kriteriums. Es interessieren weder Motiv des Akteurs, noch eventuell mit der altruistischen Handlung einhergehende kognitive oder emotionale Prozesse. Daher findet sich in der Literatur analog zur Definition des biologischen Altruismus auch die Bezeichnung des behavioralen Altruismus (siehe z.B. Fehr & Fischbacher, 2003). Anders verhält es sich beim Konzept des psychologischen Altruismus. Hierbei spezifiziert ein entscheidender Zusatz die Definition (siehe oben) darin, dass die altruistische Handlung nicht durch psychologische Vorteile des Akteurs motiviert sein darf. Im Kontext des psychologischen Altruismus wird daher auch vom „reinen“ oder „wahren“ Altruismus gesprochen (Batson, Fultz & Schoenrade, 1987; Batson, 1991).

Hilft eine Person ehrenamtlich in einer wohltätigen Einrichtung aus, so ist diese Tätigkeit zunächst als altruistisch nach der biologischen Definition zu bezeichnen. Sie opfert unentgeltlich

Zeit und Energie und nimmt somit eigene Kosten in Kauf, um anderen Menschen etwas Gutes zu tun. Wenn die Person das jedoch tut, weil sie von der Handlung ein Gefühl der Erfüllung erfährt oder diese ihr innere Ruhe und Ausgeglichenheit verleiht, liegt hierbei kein psychologischer Altruismus vor. Die Kriterien für biologischen Altruismus werden in diesem Beispiel hingegen nicht verletzt. Die behaviorale Definition schließt das Vorhandensein von Motiven des eigenen (psychologischen) Vorteils theoretisch nicht aus und bildet somit einen bedeutenden Ansatzpunkt, um das Überleben von altruistischen Verhaltensweisen im Laufe der Evolution zu erklären.

1.1.2 Altruismus in der Evolutionspsychologie

Der britische Naturforscher und Begründer der Evolutionstheorie Charles Darwin (1809 – 1882) entwickelte als Konglomerat seiner Erkundungen unter anderem in Südamerika, auf den Galapagos-Inseln, in Neuseeland und Australien seine Deszendenztheorie (1859). Diese beinhaltet die Annahme einer gemeinsamen Abstammung aller Arten und deren allmähliche Veränderung im Laufe der Zeit, die Evolution der Arten (Darwin, 1859). Dabei postulierte Darwin (1859) weiterhin, dass die am besten an ihre Umwelt angepassten Arten auch die besten Überlebenschancen aufweisen werden und deklarierte diese natürliche Selektion, gekennzeichnet durch einen stetigen Konkurrenzkampf unter den Individuen, als einen der wichtigsten Mechanismen der Evolution.

Auf den Prinzipien der Evolutionstheorie basierend wird nun im Rahmen der Evolutionspsychologie versucht das Sozialverhalten des Menschen mit genetischen Faktoren zu erklären, die im Laufe der Zeit nach den Grundgesetzen jener natürlichen Auslese entstanden sind (Buss, 2005). Werden die allgemeine Definition von Altruismus (siehe 1.1) und die Grundannahmen der Evolutionstheorie zusammen betrachtet, so ergibt sich an dieser Stelle zunächst ein Widerspruch. Ein altruistisch handelnder Mensch verhält sich scheinbar genau entgegen der wichtigsten Prinzipien der Evolutionstheorie (McAndrew, 2002). Auf einen ersten Blick möchte man meinen, dass jene „Altruismus-Gene“ des Menschen genau eine Gruppe solcher Gene bilden, welche die Überlebenschancen herabsetzen, die Chancen für Nachwuchs reduzieren und somit weniger wahrscheinlich weitergegeben werden. Führt man diesen Gedanken weiter, sollte altruistisches Verhalten im Laufe der menschlichen Evolution bereits gänzlich verschwunden sein und zwar zu Gunsten jenes Verhaltens, welches den eigenen Vorteil sichert.

Doch altruistische Verhaltensweisen haben dem Selektionsdruck der menschlichen Evolution standgehalten, ja wurden sogar von der natürlichen Auslese gefördert (Hamilton, 1964) und dies nicht zuletzt, weil ein Altruist vielleicht doch nicht so altruistisch motiviert handelt, wie es auf den ersten Blick scheinen mag (Boone, 1998).

1.1.3 Ultimate Erklärungsansätze für altruistisches Verhalten

Die folgenden Unterabschnitte beschäftigen sich mit ultimaten Ursachen für Altruismus und stellen Erklärungsmodelle vor, welche mögliche Gründe für Verhalten im evolutionstheoretischen Zusammenhang betrachten. Mit diesen Modellen wird der Frage nachgegangen, warum sich altruistische Verhaltensweisen im Laufe der Evolution gegenüber dem Selektionsdruck anderer Verhaltensweisen durchsetzen konnten. Eine entscheidende Rolle hierbei spielt der Grad der Anpassung eines Verhaltens an seine Umwelt und des daraus hervorgehenden Selektionsvorteils für das Individuum (Tinbergen, 1963). Im evolutionstheoretischen Kontext wird dieser Vorteil als gesteigerte Fitness eines Individuums gegenüber anderen bezeichnet (Darwin, 1859).

Zunächst wird das Modell der Verwandtenselektion betrachtet; im Anschluss folgen Erklärungsmodelle für altruistisches Verhalten unter genetisch nicht verwandten Individuen.

1.1.3.1 Altruistisches Verhalten unter verwandten Individuen

Das Motiv von altruistischem Verhalten unter Familienmitgliedern scheint gut vereinbar mit den Grundsätzen der Evolutionspsychologie. Selbst wenn der Akteur durch die altruistische Handlung zunächst eigene Kosten in Kauf nehmen muss, vielleicht sogar das eigene Leben riskiert, so trägt er langfristig dennoch zur Erhaltung des eigenen Genpools bei (Hamilton, 1964). Dieses Konzept der Verwandtenselektion, also von Verhalten, das einen genetisch Verwandten begünstigt, kommt insbesondere immer dann zum Tragen, wenn es sich um lebensbedrohliche Situationen handelt (Sime, 1983). Mit dem Überleben eines verwandten Menschen erhöhen sich schließlich auch die Chancen, dass die eigenen Gene in zukünftigen Generationen fortleben werden. Dies gilt speziell für nahe Verwandte, wie zum Beispiel Geschwister, Eltern oder Kinder, denn diese weisen einen besonders hohen Anteil gemeinsamer Gene auf. Da dieser Anteil mit zunehmendem Grad der Verwandtschaft abnimmt, sinkt ebenso der Effekt der bevorzugten Auswahl von Verwandten, wenn es um den Einsatz altruistischer Verhaltensweisen geht (siehe Burnstein, Crandall & Kitayama, 1994).

1.1.3.2 Altruistisches Verhalten unter nicht verwandten Individuen

Im Gegensatz zu altruistischem Verhalten unter verwandten Individuen, welches auch bei einzelnen Arten im Tierreich, wie zum Beispiel bei Hummeln, Bienen oder Ameisen, zu finden ist, gilt das häufige Auftreten von altruistischem Verhalten unter nicht verwandten Individuen als einzigartig für die menschliche Spezies (Boyed & Richerson, 2004). Die modernen Gesellschaftsformen verschiedener Kulturen basieren auf einem System von Arbeitsteilung und Kooperation in großen Gruppen von genetisch nicht verwandten Individuen. Doch auch bereits frühe, einfache Formen von Gemeinschaften wiesen in der Regel ein funktionierendes Netzwerk kooperativer Verhaltensweisen auf, sei es bei der Jagd, der Nahrungsaufteilung oder bei der Kriegsführung und dies ganz ohne richtungsweisende Judikative, Exekutive und Legislative (Hill, 2003; Kaplan, Hill, Lancaster & Hurtado, 2000).

Evolutionstheoretisch betrachtet scheint die Fitness von Mensch und Tier langfristig zu steigen, wenn gegenseitige Bereitschaft zu kooperativem Verhalten von allen Gruppenmitgliedern gezeigt wird (Fehr et al., 2003). Doch während altruistisches Verhalten unter Tieren weitestgehend auf genetisch eng verwandte Individuen beschränkt ist, zeigen sich beim Menschen noch weitere, spezifische Formen von altruistischem Verhalten. Erklärungsmodelle hierfür stützen sich vor allem auf die Prinzipien der direkten, indirekten und der starken Reziprozität.

Direkte Reziprozität Im Volksmund heißt es: „Wie du mir, so ich dir!“. Ein Motiv, das den Anspruch erhebt altruistische Verhaltensweisen gegenüber nicht verwandten oder fremden Personen erklären zu können, lässt sich in ähnliche Worte fassen: „Wie ich dir, so (erwarte ich) du mir!“. Das Anfang der 70er Jahre von Trivers (1971) entwickelte Konzept des reziproken Altruismus besagt, dass Menschen sich altruistisch verhalten, weil sie davon ausgehen bei gegebener Situation ebenso altruistisches Verhalten vom Gegenüber zu erhalten. Das Folgen von Kooperation auf Kooperation und umgekehrt das Ausbleiben kooperativer Verhaltensweisen, falls vorausgehend nicht kooperiert wurde, beschreibt das Prinzip der *Tit-for-Tat-Strategie* und konnte empirisch nachgewiesen werden (Axelrod & Hamilton, 1981; siehe Abbildung 1a).

Der Nutzen aus der reziprok altruistischen Interaktion ist in der Regel größer als die erbrachten Kosten. Daher bringt der Einsatz von reziprok altruistischem Verhalten langfristig einen Fitnessvorteil für beide Interaktionspartner und kann auch in den Erklärungsrahmen der Evolutionspsychologie integriert werden. Nach dem Prinzip der natürlichen Auslese scheinen über Generationen hinweg solche Individuen einen Selektionsvorteil besessen zu haben, welche den

kooperativen Austausch mit ihrem unmittelbaren sozialen Umfeld nach dem Prinzip der Reziprozität entwickelt haben. Sie investierten somit nicht „unnötig“ eigene Ressourcen in eine selbstlose Hilfe für andere, sondern nur dann, wenn sie zukünftig selbst hiervon profitieren konnten und sich der Kosteneinsatz somit auszahlte (Trivers, 1971).

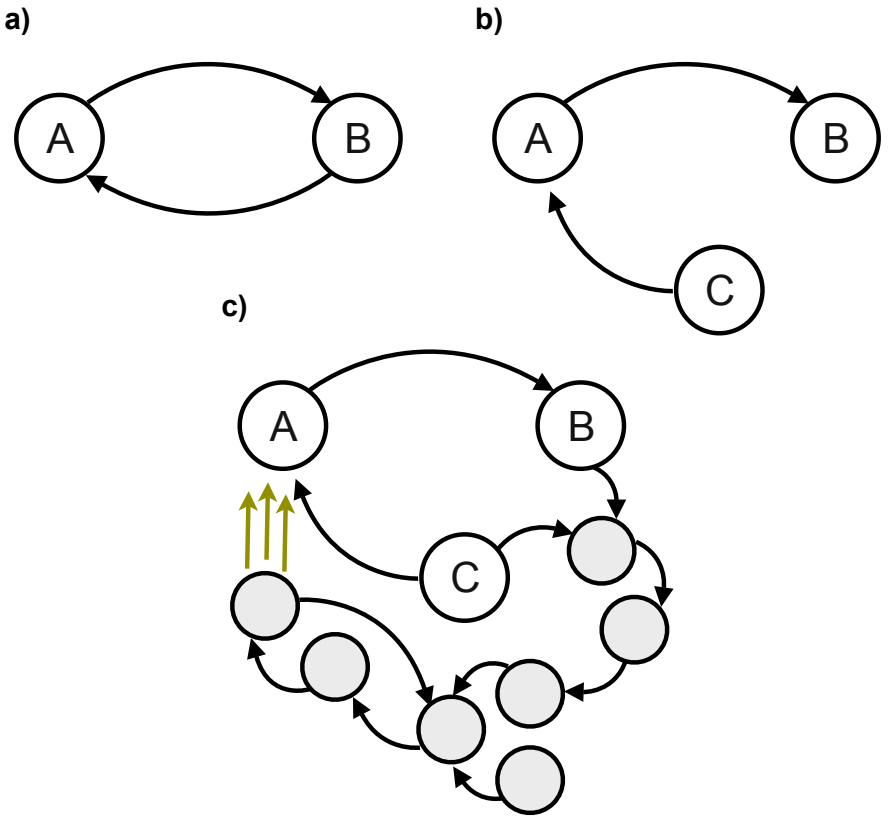

Abbildung 1: Direkte und indirekte Reziprozität. a) Nach dem Prinzip der direkten Reziprozität hilft Person A Person B und Person B Person A in wiederkehrenden Interaktionen. b) Nach dem Prinzip der indirekten Reziprozität hilft Person A Person B auch in einmaliger Handlungsaktion. Hierfür erhält Person A Hilfe von Person C, welche das Geschehen beobachtet hat und Person A als einen guten Menschen einschätzt, der es verdient selbst Hilfe zu erhalten. c) Der gute Ruf von Person A kann über Person C und Person B weiter an andere Mitglieder der Gruppe kommuniziert werden, so dass Person A auch durch die indirekt reziproke Hilfeleistung derer profitiert, welche beim eigentlichen Geschehen überhaupt nicht anwesend waren (Nowak & Sigmund, 2005). In allen drei Fällen zieht Person A aus ihrem ersten Handlungsschritt zu Gunsten von Person B, direkt (a) oder indirekt (b, c) einen eigenen langfristigen Nutzen (Abbildung in Anlehnung an Nowak & Sigmund, 2005).

Handelt jedoch einer der Partner entgegen der Reziprozitätsnorm, muss er damit rechnen, dass in Zukunft das System der Kooperation nach der Tit-for-Tat-Strategie zusammenbricht. Dies sei mit Hilfe eines einfachen Beispiels illustriert:

Wir stellen uns vor, es leben zwei Bauern auf ihrem Gut nicht weit von einander, jedoch fernab von jeglicher Zivilisation. Bauer A besitzt ein Stück Grasland, auf welchem zahlreiche Kühe weiden. Auf dem Gut von Bauer B hingegen wird schon seit Generationen Getreide angebaut. Jedes Jahr im Herbst bringt Bauer A ein paar Kästen seiner abgefüllten Flaschen Milch zu Bauer B. Gibt dieser daraufhin auch ein paar Säcke seines Getreidevorrats ab, handelt Bauer B reziprok altruistisch. Nimmt er die Milch jedoch an, ohne etwas zu erwidern, muss er damit rechnen, nächstes Jahr im Herbst auch keine Milch von Bauern A mehr zu bekommen. Beiden Bauern müsste an einer gegenseitigen Kooperation gelegen sein, denn die Kosten sind aufgrund des Überflusses an jeweils eigenem Rohstoff begrenzt, doch der Nutzen, einen Teil vom kostbaren Rohstoff des anderen zu erhalten, ist groß.

Das Wissen um den Verlust des eigenen langfristigen Vorteils bei Bruch der Reziprozitätsnorm kann eine Erklärung für das Auftreten von direkt reziprok altruistischem Verhalten bilden, vorausgesetzt es besteht eine hinreichend hohe Wahrscheinlichkeit zukünftig erneut mit einem entsprechendem Partner zu interagieren (Friedman, 1971; Gouldner, 1960). Jedoch bleibt nach dem Konzept der direkten Reziprozität unklar, warum sich Menschen auch in anonymen beziehungsweise einmaligen Handlungsaktionen altruistisch verhalten.

Indirekte Reziprozität Im vorangegangenen Unterabschnitt wurde darauf verwiesen, dass das System der direkten Reziprozität auf der Einhaltung einer klar definierten Norm durch beide Interaktionspartner basiert. Wird diese von einem der Partner gebrochen, besteht für den anderen ein erhöhtes Risiko mehr Kosten zu investieren als Nutzen zu erlangen, wenn er weiterhin kooperativ in der Interaktion handelt. Eine Erklärung dafür, warum Menschen auch dann altruistisch handeln, wenn eine Interaktion nach dem Prinzip der direkten Reziprozität auf großer Unsicherheit basiert beziehungsweise erst gar keine Möglichkeit hierfür besteht, bildet das Prinzip der indirekten Reziprozität. „Wie ich dir, so (erwarte ich) andere mir“ könnte die Theorie in Kurzform lauten. In der Tat verhalten sich Menschen eher kooperativ gegenüber solchen Mitmenschen, von denen sie wissen, dass diese auch kooperativ gegenüber anderen gehandelt haben (Lotem, Fishman & Stone, 1999; Nowak & Sigmund, 1998). Das Prinzip der indirekten Rezi-

prozität basiert auf dem Ruf des „guten“, hilfsbereiten Menschen und ist somit unabhängig von der direkten Reaktion des Interaktionspartners (siehe Abbildung 1b). Der sich daraus ergebende langfristige Nutzen für die altruistisch handelnde Person erklärt dabei den Kosteneinsatz, welcher zunächst erbracht werden muss, um einen solchen Ruf zu etablieren. Eine Erweiterung des oben angebrachten Beispiels kann auch zur Verdeutlichung der indirekten Reziprozität beitragen:

Angenommen es befinden sich neben dem Gut A und Gut B noch eine Anzahl von weiteren Gütern in dem sonst isolieren Gebiet. Die ursprünglichen Eigentümer von Gut A haben ihr Gut zur Miete abgegeben. Nun kommt jedes Jahr im Herbst ein anderer Bauer A, um die Kühe zu melken. Jedes Jahr gibt auch ein anderer Bauer A etwas von seinem Milchvorrat an den Nachbarn Bauern B ab und jedes Mal kann Bauer B diese Gabe mit einer entsprechenden Menge Getreide erwidern oder auch nicht. Nach dem Prinzip der direkten Reziprozität wäre es für Bauer B ökonomischer, jedes Jahr die Milch anzunehmen und das gesamte Getreide für sich zu behalten. Doch Bauer B weiß, dass sein Verhalten von den anderen Gutsbesitzern beobachtet wird. Auf deren Gunst ist Bauer B angewiesen, da diese im Besitz von anderen, wertvollen Rohstoffen wie Eier oder Wolle sind. Daher wird Bauer B auch immer etwas Getreide für die jährlich wechselnden Bauern A als Gegenleistung für die Milch bereithalten. Bauer B wird somit für sein kooperatives Verhalten unter den anderen Bauern bekannt werden und einen guten Ruf erhalten. Sie werden ihm daher vertrauen und zu gegebenem Zeitpunkt auch etwas von ihren Rohstoffen anbieten.

Voraussetzungen für das Funktionieren einer solchen Theorie bestehen wie auch im Beispiel illustriert darin, dass andere Personen überhaupt Notiz von der altruistischen Handlung nehmen, diese im Gedächtnis behalten und bestenfalls an andere weitergeben (siehe Abbildung 1c). Daher weißt auch dieses Modell argumentative Lücken auf und kann nicht erklären, warum sich Menschen auch in anonymen Handlungsaktionen ohne Publikum altruistisch verhalten (Fehr & Fischbacher, 2003). Weiterhin kann davon ausgegangen werden, dass das Aufbauen und Erhalten einer entsprechenden Reputation in der Regel nur in Gruppen von begrenztem Umfang möglich ist. In sehr großen, bezüglich sozialer Richtlinien nicht klar definierten Gruppen steigt die Wahrscheinlichkeit, dass die Norm darüber was ein „guter“ Ruf ist unter den Gruppenmitgliedern divergiert und es sinkt die Wahrscheinlichkeit, dass Handlungen einzelner Individuen explizit und nachhaltig wahrgenommen werden (Nowak & Sigmund, 2005).

Direkte und Indirekte Reziprozität – Zusammenfassung Sowohl ein reziprok altruistischer Akteur als auch ein auf die eigene Reputation bedachter Akteur handelt deshalb kooperativ, weil er sich langfristig einen eigenen Nutzen von der kurzfristigen Investition verspricht. Lässt sich dieser Vorteil nicht antizipieren, wird nach den Theorien der direkten und indirekten Reziprozität auch kein kooperatives Verhalten gezeigt werden. Das Motiv des Handelnden besteht in beiden Fällen darin, den eigenen Nutzen zu maximieren beziehungsweise die eigene Fitness langfristig zu erhöhen. Daher stellt sich die Frage, ob es gerechtfertigt ist, jenes Verhalten überhaupt noch als altruistisch zu bezeichnen. Manche Autoren gehen sogar soweit, die Begrifflichkeiten *reziproker Altruist* und *egoistisch* synonym zu verwenden (Fehr et al., 2003).

Starke Reziprozität Eine Erweiterung der Theorien von direkter und indirekter Reziprozität bildet das Konzept der starken Reziprozität zur Erklärung von altruistischem Verhalten. Eine Person handelt dann stark reziprok, wenn sie gewillt ist eigene Ressourcen zu investieren, um faires, normkonformes Verhalten anderer zu belohnen (starke positive Reziprozität) und unfaires, normverletzendes Verhalten anderer zu bestrafen (starke negative Reziprozität – Fehr et al., 2003; Fehr & Fischbacher, 2003; siehe Abbildung 2). Wann eine Handlung als fair beziehungsweise als unfair beurteilt wird, hängt zum einen von ihren Konsequenzen ab und zum anderen von den der Handlung zu Grunde liegenden Intentionen des Akteurs (Falk & Fischbacher, 2006). Dieser Sachbestand soll mit Hilfe des bereits eingeführten Beispiels in abgewandelter Form näher erläutert werden:

Wie jeden Herbst kommt ein neuer Bauer A, um die Kühe zu melken. Gehen wir zuerst davon aus, der Milchbauer A hätte in diesem Jahr nur einen kleinen Vorrat an Milch, genau eine große und eine kleine Flasche voll, erwirtschaften können, da die Kuhherde krank gewesen ist. Nehmen wir weiterhin an, dass es dieses Mal Bauer B ist, welcher zuerst zu Bauer A kommt, um diesem einen Sack Getreide zu bringen. Erwidert Bauer A den Getreidesack mit der kleinen Flasche Milch, so wird Bauer B dies in Anbetracht der Notlage von Bauer A durchaus noch als fair beurteilen und nach dem Prinzip der starken Reziprozität zum Beispiel dadurch belohnen, dass er dem Bauern B zusätzlich noch einen zweiten Sack Getreide schenkt. Die altruistische Belohnung kann aber auch durch Dritte, wie die anderen Bauern erfolgen. Diese beobachten das normkonforme Verhalten von Bauer A und geben diesem ebenfalls etwas von ihren Rohstoffen ab. Gehen wir nun davon aus, Bauer A hätte ein ertragreiches Jahr gehabt und mehrere große

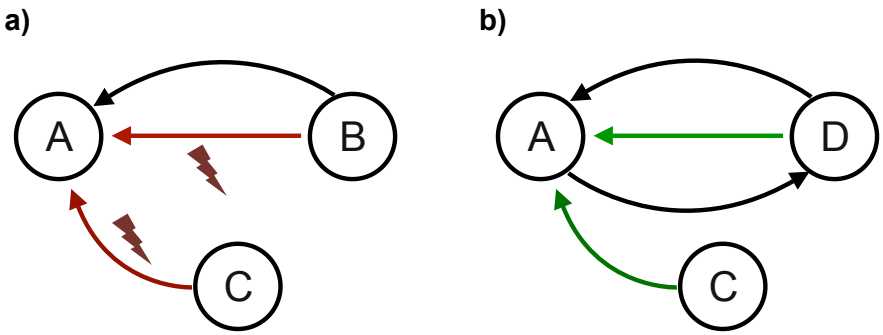

Abbildung 2: Die zwei Komponenten der starken Reziprozität: Altruistische Bestrafung (a) und altruistische Belohnung (b). Person A wird von Person C und/oder Person B altruistisch bestraft (roter Pfeil), nachdem sie das kooperative Verhalten von Person B nicht erwidert und somit eine soziale Norm gebrochen hat (a) und von Person B und/oder Person C belohnt (grüner Pfeil), nachdem diese das kooperative Verhalten von Person B der Norm entsprechend erwidert hat (b). Person C und/oder Person B ziehen weder aus der altruistischen Bestrafung noch aus der altruistischen Belohnung einen eigenen Nutzen.

und kleine Flaschen mit Milch abfüllen können. Gibt Bauer A für den großen Getreidesack auch unter dieser Gegebenheit nur eine kleine Flasche Milch an Bauern B, gilt dieses Verhalten als unfair und wird nach dem Prinzip der starken Reziprozität altruistisch bestraft. So könnten Bauer B als auch die anderen Gutsbesitzer die Bitte von Bauer A abschlagen, ihm beim Aufladen der Flaschen für den Transport zu helfen, obwohl sowohl Bauer B als auch die anderen sich damit eigentlich gern wie üblich etwas Kleingeld dazu verdient hätten.

Die zwei Varianten des Beispiels verdeutlichen, dass eine Handlung trotz gleicher Konsequenz (eine kleine Flasche Milch) als normkonform oder aber normverletzend wahrgenommen werden kann (Falk & Fischbacher, 2006). Auch wird im Beispiel bereits das entscheidende Merkmal der starken Reziprozität insbesondere in Abgrenzung zur direkten und indirekten Reziprozität veranschaulicht. Dieses besteht darin, dass eine stark reziprok handelnde Person dazu bereit ist auch dann eigene Kosten in Kauf zu nehmen, wenn sich diese Investition weder kurz- noch langfristig für die eigene Person materiell auszahlen wird. Dies trifft für die stark reziprok agierenden Bauern im Beispiel zu, da sie wissen, dass die Interaktion mit Bauer A nur einmalig stattfinden wird (Fehr et al., 2003).

Die Theorie der starken Reziprozität als ein Vertreter der ultimaten Erklärungsansätze für altruistisches Verhalten wurde als Theorie konzipiert, nach welcher das Fortbestehen von Al-

truismus unter dem Gesichtspunkt des sich hieraus ergebenden Selektionsvorteils argumentiert werden kann. Das evolutionstheoretische Modell von Gintis (2000) bietet hierfür zunächst gute Ansatzpunkte. Gintis (2000) argumentiert, dass besonders in Situationen von akuter Gefahr, während eines Krieges oder bei Umweltkatastrophen, eine Gemeinschaft darauf angewiesen ist, dass alle Mitglieder zusammen arbeiten. Um den Einsatz jedes einzelnen Individuums systematisch zu verstärken, wird kooperatives Verhalten belohnt sowie davon abweichendes, egoistisch motiviertes Verhalten bestraft und somit von vornherein besser unterdrückt. Denn bereits das Wissen um die Möglichkeit für nicht kooperatives Verhalten bestraft zu werden steigert den Grad an Kooperation erheblich (Fehr & Gächter, 2002). Die Individuen von Gruppen mit einer bestimmten Anzahl von stark reziprok handelnden Mitgliedern haben somit einen Fitnessvorteil gegenüber den Individuen anderer Gruppen, da langfristig jedes einzelne Mitglied von der gemeinschaftlichen Kooperation profitiert. Unter Bedrohung ist dieser Vorteil für die Gruppe größer als die erbrachten Kosten der altruistisch bestrafenden Gruppenmitglieder (Gintis, 2000).

Dennoch muss auch an dieser Stelle auf inhaltliche Lücken verwiesen werden. Zum einen funktioniert der Aspekt des Selektionsvorteils nur auf Gruppenebene. Innerhalb der Gruppe haben solche Mitglieder, welche zwar kooperieren, jedoch nicht bestrafen, einen Fitnessvorteil gegenüber stark reziprok handelnden Mitgliedern. Somit kann das Modell nicht erklären, wieso Verhaltensweisen nach dem Prinzip der starken Reziprozität dem Selektionsdruck, ausgehend von rein kooperativen Verhaltensweisen, im Laufe der Evolution standhalten konnten. Denn würden alle Individuen einer Gruppe kooperieren, gäbe es auch keinen Anlass zur Bestrafung und somit keine Benachteiligung für einzelne Mitglieder (Fehr et al., 2003). Zum anderen kann das ultimate Modell nicht erklären, warum Menschen auch außerhalb eines unmittelbaren Gruppenkontextes dazu gewillt sind stark reziprok zu handeln (Fehr & Gächter, 2002). Proximate Modelle scheinen hier Abhilfe zu bieten. Eine ausführliche Betrachtung derer erfolgt im nächsten Abschnitt, welcher sich im Detail dem Phänomen der altruistischen Bestrafung widmet.

1.2 Altruistische Bestrafung

Altruistische Bestrafung, die Neigung eines Individuums normverletzendes Verhalten zu bestrafen, und altruistische Belohnung, die Neigung eines Individuums normkonformes Verhalten zu belohnen, bilden die zwei Bestandteile für eine Form von altruistischem Verhalten, der starken Reziprozität.

In den folgenden Ausführungen wird sich der Fokus auf einen Teil der starken Reziprozität und zugleich Kernstück der vorliegenden Arbeit richten, auf das Phänomen der altruistischen Bestrafung. Zuerst werden Nutzen und Kosten der altruistischen Bestrafung noch einmal detailliert im Kontext der Erhaltung von sozialen Normen erläutert. Im Anschluss wird unter Bezugnahme auf proximate Modelle versucht, den Erklärungsrahmen für das Phänomen der altruistischen Bestrafung zu erweitern.

1.2.1 Nutzen und Kosten von altruistischer Bestrafung

Kooperation gilt als eine soziale Norm und bildet ein wichtiges Element für effizientes Zusammenleben vieler Individuen in einer Gruppe. Um das Risiko zu vermindern, dass einzelne Individuen die Vorteile einer durch Kooperation geprägten Gemeinschaft nutzen, ohne selbst zu kooperieren, wird nicht kooperatives Verhalten bestraft. Der Vorteil dieses Free Ridings geht durch die Bestrafung verloren und lohnt sich nicht mehr. Somit führt allein die Antizipation der Bestrafung bei Normbruch dazu, dass in der sozialen Interaktion verstärkt kooperativ gehandelt wird. Trifft nun ein kooperatives Mitglied der Gruppe auf ein ursprünglich nicht kooperatives Mitglied, welches jedoch aufgrund früherer Bestrafungen oder wegen des Wissens über die Möglichkeit hierfür bestraft werden zu können nun auch kooperativ handelt, wird das prosoziale Element der altruistischen Bestrafung deutlich (Fehr & Gächter, 2002). Die Bestrafung kann dabei direkt durch die betroffene Person geschehen (First Person) oder indirekt von Dritten ausgehen (Third Party). Eine Person A wird für ein normverletzendes Verhalten in der Interaktion mit Person B also entweder von der Person B selbst oder von einer beobachtenden Person C bestraft. Dies führt dazu, dass sich Person A beim Aufeinandertreffen mit einer anderen Person D mit einer höheren Wahrscheinlichkeit normkonform verhalten wird, Person D also den Nutzen aus der altruistischen Bestrafung durch Person B oder Person C zieht. Um es mit den Worten von Fehr und Gächter zusammenzufassen (2002, Seite 137): „In this sense, punishment is altruistic. In the presence of altruistic punishers, even purely selfish subjects have a reason to cooperate in the punishment treatment“ (siehe Abbildung 3).

Doch worin bestehen die Kosten des altruistisch Bestrafenden? Diese können sowohl quantitativ als auch qualitativ in verschiedenster Weise ausfallen. So kann eine stark reziproke Person einerseits dazu gewillt sein, eigene materielle Güter für die Bestrafung zu investieren. Hier ist der

Kosteneinsatz objektiv recht gut ersichtlich. In alltäglichen Situationen, in denen altruistische Bestrafung auftritt, bedarf es andererseits häufiger eines subtileren Einsatzes von nicht materiellen Ressourcen, wie Zeit und Kraft. Im Rahmen von Internetauktionen ist es zum Beispiel möglich ein öffentliches Feedback der eigenen Zufriedenheit mit einem entsprechenden Händler in schriftlicher Form zu hinterlassen.

Nehmen wir an, ein Käufer (B) fühlt sich ungerecht behandelt, da ihm für sein Geld eine höhere Qualität des Produktes oder ein besserer Service zugestanden hätte. Macht er sich nun die Mühe einen ausführlichen, negativen Meinungsbericht zu verfassen und diesen auf die Internetseite des Verkäufers (A) zu stellen, obwohl es sich für ihn von Anfang an um eine einmalige Aktion gehandelt hat, dann bestraft er altruistisch. Der Händler weiß, dass das kritische Feedbackschreiben auch in Zukunft für potentiell neue Käufer sichtbar sein wird. Da er an weiteren Geschäften interessiert ist, wird er alles geben, um das negative Schreiben durch positive auszugleichen und die Qualität seiner Produkte sowie den Service steigern oder den Verkaufspreis senken. Somit profitieren andere Käufer (D) von den Kosten, welche das Schreiben von Käufer B mit sich gebracht hat.

Der Käufer B bestraft Händler A altruistisch, weil er sich von diesem „unfair“ behandelt gefühlt hat. Er hat sich sehr darüber „geärgert“, insbesondere da sein Verhalten durch eine sehr

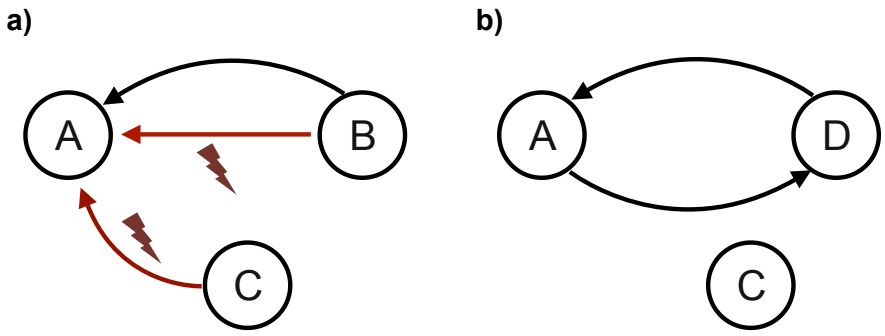

Abbildung 3: Der Einfluss von altruistischer Bestrafung auf das Entstehen und die Erhaltung von Kooperation. a) Person A wird von Person B und/oder Person C bestraft, nachdem Person A die Norm von gegenseitiger Kooperation gebrochen hat. b) Person A verhält sich nun in erneuter Interaktion mit einer anderen Person D kooperativ, da Person A gelernt hat, dass sich nicht kooperatives Verhalten nicht lohnt, wenn es bestraft wird. Person B und/oder Person C bestrafen somit altruistisch, da die Bestrafung einer anderen Person (D) einen Nutzen bringt und für die Bestrafenden selbst nur Kosten bedeuten.

hohe „Sensibilität für Ungerechtigkeit“ aus der Perspektive des Beobachters gekennzeichnet ist. Auch die Vorstellung, dass andere Kunden in dieselbe Situation geraten könnten, bewegt Käufer A zu dem Schreiben. Er kann sehr gut „nachempfinden“, wie es diesen ergehen würde und ist gewissermaßen „befriedigt“ etwas gegen diese Ungerechtigkeit tun zu können. Fairnessnorm, Emotionen, Ungerechtigkeitssensibilität und Empathie stellen Beispiele für proximate Erklärungskonzepte für altruistische Bestrafung dar und werden in den folgenden Unterabschnitten näher erläutert.

1.2.2 Proximate Erklärungsansätze für altruistische Bestrafung

Proximate Modelle beschreiben Auslöser für Verhalten, welche mit der aktuellen Handlungssituation unmittelbar einhergehen. Diese können sowohl aus dem Individuum selbst hervorgehen, wie psychische, physiologische und neurochemische Faktoren, oder von der Umwelt induziert sein. Ein Beispiel für letzteres stellen die aktuellen sozialen Bedingungen und Normen dar, unter welchen das Individuum handelt. Aber auch Vorerfahrungen und Entwicklungs- oder Reifungsprozesse fallen in die Kategorie der proximaten Ursachen für Verhalten (siehe Tinbergen, 1963).

1.2.2.1 Soziale Normen

Jede menschliche Gemeinschaft ist durch bestimmte Normen charakterisiert (Fehr & Fischbacher, 2004a). Soziale Normen bilden Richtlinien für Verhaltens- und Wertestandards, welche in einer Gruppe als anerkannt gelten und häufig in Form von ungeschriebenen Vorschriften oder impliziten Erwartungen vorliegen (Elster, 1989).

Ein Beispiel für eine soziale Norm bildet die Fairnessnorm. Mittels dieser werden unter anderem solche Verhältnisse definiert, die bei der Aufteilung eines bestimmten Guts von x Einheiten zwischen zwei Personen als fair gelten. Sind beide gleich berechtigt, das heißt keiner der Partner hat zuvor mehr oder weniger als der andere für seinen Anteil investiert (*Input*), gilt die Aufteilung dann als fair, wenn beide Personen auch den gleichen Anteil, also x/2 Einheiten, erhalten (*Output*). In Bezug auf die Fairnessnorm sollen sich Input und Output die Waage halten; es soll *Equity* bestehen. Abweichungen hiervon erzeugen *Distress* in Form eines negativen Spannungsfeldes bei der betroffenen Person (Adams, 1965; Homans, 1961; Walster, Berscheid & Walster, 1973). Das davon ausgehende Bedürfnis nach Entladung und Ausgleich könnte eine Erklärung

dafür bilden, warum Gruppenmitglieder, welche entsprechende Vorschrift brechen, bestraft werden. Tritt eine Person neu in eine Gruppe mit bereits bestehenden sozialen Normen ein, wird sie diese von den anderen Mitgliedern lernen. Das Beobachten oder Erfahren der negativen Konsequenzen beim Verletzen einer Norm unterstützt den Lernprozess und trägt zugleich stabilisierend zur Normerhaltung bei (Fehr & Fischbacher, 2004a).

An dieser Stelle ergeben sich mindestens zwei mögliche Motive zur Begründung des Auftretens von altruistischer Bestrafung.

Altruistische Bestrafung als eine soziale Norm Das erste Motiv besteht in der Existenz von sozialen Normen an sich. In Gruppen herrschen soziale Normen, ihre Mitglieder lernen sich diesen Normen nach adäquat zu verhalten (siehe oben) und eine dieser Normen besteht in der altruistischen Bestrafung von unfairem Verhalten.

Dieser Lernprozess, auch unter in Kaufnahme von eigenen Kosten zu bestrafen, könnte wiederum dadurch verstärkt werden, dass weiterhin die Norm besteht, ebenfalls solche Individuen altruistisch zu bestrafen, welche einen wahrgenommenen Normbruch nicht altruistisch bestrafen. Dieses „Punishment of non-punishers“ (Fehr & Fischbacher, 2003, Seite 790) würde einen, allerdings ultimaten, Erklärungsansatz dafür bilden, warum soziale Normen auch in großen Gruppen ($n > 16$) aufrecht erhalten bleiben (Boyd, Gintis, Bowles & Richerson, 2003).

Altruistische Bestrafung zur Erhaltung von sozialen Normen Das zweite Motiv im Rahmen von sozialen Normen bezieht sich hingegen auf deren Erhaltung. Da das Vorherrschen von Normen, wie bereits erwähnt, ein existenzielles Merkmal einer Gesellschaft bildet und das gemeinsame Zusammenleben der Individuen stützt und erleichtert, kann angenommen werden, dass es im Interesse der Gruppe ist, jene Normen zu schützen. Hierfür bedarf es in der Regel jedoch des Einsatzes von Ressourcen, um solche Mitglieder der Gruppe zu bestrafen, welche sich normverletzend verhalten. Wenn nur einige Individuen sozialen Normbruch bestrafen, jedoch auch andere, insbesondere kooperierende und nicht bestrafende Individuen von dieser Bestrafung profitieren, wird von altruistischer Bestrafung gesprochen (Fehr & Fischbacher, 2004b; siehe 1.2.1).

1.2.2.2 Emotionen

Menschen reagieren emotional in Situationen, welche als persönlich bedeutsam wahrgenommen werden (Frijda, 1988). Diese Reaktion kann von positiver oder negativer Valenz geprägt sein und

umfasst ein komplexes Muster von intraindividuellen Vorgängen. Das physiologische Erregungsniveau verändert sich, spezifische Gefühle werden wahrgenommen und für die Emotion typische Verhaltensreaktionen und -ausdrücke gezeigt; kognitive Prozesse können verzerrt werden (Elster, 1998). Letzteres kann je nach Intensität der Emotion auch dazu führen, dass eine rationale Kosten- und Nutzenabwägung eingeschränkt wird. Es konnte gezeigt werden, dass eine stark emotional involvierte Person ihr Handeln eher von der Emotion als von der Kognition leiten lässt im Vergleich zu einer weniger emotional involvierten Person (Elster, 1998; Frank, 1988). Mit Bezug auf den nach materiellem Eigennutz strebenden *Homo oeconomicus* (siehe 1.3.1) wird die emotionale Person irrational und in diesem Kontext vielleicht altruistisch handeln. Denn ein altruistischer Akt ist ökonomisch betrachtet immer ein irrationaler Akt (Fehr & Fischbacher, 2003).

Die folgenden Ausführungen legen dar, inwiefern Emotionen sowohl von negativer als auch von positiver Valenz die Entscheidung zu altruistischer Bestrafung beeinflussen könnten.

Emotionen negativer Valenz In Unterabschnitt 1.2.2.1 wurde anhand von zwei theoretischen Gesichtspunkten versucht zu begründen, warum die Wahrnehmung einer gebrochenen Fairnessnorm altruistisch bestraft wird. Der Zusammenhang zwischen sozialem Normbruch und altruistischer Bestrafung lässt sich jedoch noch weiter spezifizieren. Betrachtet man die gebrochene Norm als ein persönlich bedeutsames Ereignis, dann kann die daraus resultierende Emotion als eine mediierende Variable für die Bereitschaft zu altruistischer Bestrafung interpretiert werden. Verschiedene Autoren verweisen darauf, dass mit der Wahrnehmung von Free Riding stark negative Emotionen, wie zum Beispiel Ärger, einhergehen, welche wiederum den Willen erzeugen können, die nicht kooperative Person zu bestrafen (Elster, 1989; Fehr & Gächter, 2002; Frank, 1988). Elster (1989) postuliert weiter, dass eine stark negativ emotional involvierte Person gegebenenfalls auch Verhalten zeigt, welches dem Eigennutz undienlich ist. Somit ließe sich der Einsatz von eigenen Ressourcen auch in einmaligen Handlungsaktionen außerhalb eines Gruppenkontextes erklären, also dann, wenn das Motiv zur Erhaltung der eigenen Gruppennorm nicht gegeben ist.

Emotionen positiver Valenz Die Autoren de Quervain et al. (2004) argumentieren mit positiven Emotionen zur Erklärung von altruistischer Bestrafung. Unter Verwendung der *Positronenemissionstomographie* (PET, zur Methode der PET siehe z.B. Walter, 2005, Kapitel 1) konnten

sie zeigen, dass das Ausführen von altruistischer Bestrafung mit einer verstärkten neuronalen Aktivierung solcher Regionen im Gehirn einhergeht, welche mit Belohnung assoziiert sind. Bestraft eine Person aufgrund eines wahrgenommenen Normbruchs altruistisch, so kann dies als eine durch Rache motivierte Reaktion einer direkt betroffenen Person interpretiert werden. Die Genugtuung darüber, es der anderen Person „heimzahlen“ zu können, würde in diesem Fall die kostenausgleichende Belohnung darstellen. Die Bestrafung könnte jedoch auch als befriedigende Möglichkeit dafür betrachtet werden, die von einem Normbruch ausgehende Ungerechtigkeit beheben zu können. Denn geschieht die Bestrafung in einem Umfang, der qualitativ und quantitativ vergleichbar ist mit dem Vorteil, welchen sich das nicht kooperierende gegenüber dem kooperierenden Individuum erschlichen hat, herrscht nach dem Konzept der Equity Theorie (siehe 1.2.2.1) wieder Gleichheit, allerdings nur unter der Voraussetzung, dass die Kosten des altruistisch Bestrafenden ebenfalls in irgendeiner Form ausgeglichen werden. Dies kann durch Freude oder Erleichterung, also durch positive Emotionen geschehen, welche eine Person mit dem aktiven Einsatz für Gerechtigkeit assoziiert. Während eine Person das normwidrige Verhalten einer anderen Person erfährt oder beobachtet, wird bereits die Antizipation des berechtigten Eingreifens in Form von Bestrafung die positiven Emotionen herbeiführen. Somit kann auch erklärt werden, warum in der Studie von de Quervain et al. (2004) solche Probanden besonders stark bestraften, welche zuvor eine besonders hohe Aktivierung in Belohnungsarealen des Gehirns, als ein neuronales Korrelat von positiven Emotionen, aufgewiesen haben.

1.2.2.3 Empathie

Nach der Empathie-Altruismus-Hypothese (Batson, 1991) verhalten sich Menschen altruistisch, weil sie Empathie für einen anderen Menschen empfinden. Ausgehend von der *Simulation-Theory* (Gallese & Goldman, 1998) beschreibt Empathie die Fähigkeit, sich in die Lage eines anderen Menschen zu versetzen und die hiermit einhergehenden Emotionen und Ereignisse so nachzuempfinden, wie der andere sie erlebt. Somit kommt dieses Motiv insbesondere immer dann zum Tragen, wenn eine andere Person in unangenehmen, leidvollen Situationen wahrgenommen wird. In diesem Kontext entwickeln Menschen das Bedürfnis ohne Rücksicht auf eigene Interessen die Not des Gegenübers zu lindern. Daher sind sie auch bereit eigene Kosten für das Wohl des anderen zu investieren ohne selbst einen Nutzen hieraus zu ziehen. Sie zeigen damit altruistisches Verhalten nach der psychologischen Definition, handeln altruistisch nach dem „reinen“ Motiv (Batson, 1991; Batson, 1998; Batson & Moran, 1999).

Die Begründung für den Einsatz von altruistischer Bestrafung durch Dritte lässt sich nach Batsons Annahme (1991) wie folgt herleiten: Beobachtet eine Person C das ungerechte Verhalten einer Person A gegenüber einer anderen Person B, wird Person C nach der Empathie-Altruismus-Hypothese die unfaire Behandlung von Person B nachfühlen können. Sie wird das Ziel haben die Notlage von Person B, resultierend aus der Ungerechtigkeit, zu mildern. Ist nur eine indirekte Interventionsmöglichkeit über Person A gegeben, wird diese durch Person C bestraft werden. Person C wird dies tun, weil sie davon ausgehen kann, dass nun auch Person B sich besser fühlen wird, da Person A ihre gerechte Strafe (siehe 1.2.2.1 und 1.2.2.2) erhalten hat.

1.2.2.4 Ungerechtigkeitssensibilität

Für die Erklärung und zur Vorhersage von Verhalten bedarf es meist persönlicher und situativer Faktoren zugleich. Während soziale Normen ein Beispiel für den Einfluss der Umwelt und der Situation auf das Verhalten einer Person darstellen, bilden Emotionen und Empathie persönliche Einflussfaktoren. Nun reagieren Menschen selbst unter identischen situativen Umständen nicht immer gleich bezüglich der Ausprägung ihrer persönlichen Merkmalen (Asendorpf, 2005). Das Konzept der Ungerechtigkeitssensibilität beschreibt eines dieser Merkmale und zwar die interindividuellen Unterschiede in Bezug auf die Toleranz gegenüber Ungerechtigkeit und moralischem Normbruch (Schmitt, 1996).

Ungerechtigkeit kann aus verschiedenen Perspektiven wahrgenommen werden. Eine Person kann selbst den Auslöser für die Ungerechtigkeit bilden, sie kann von der Ungerechtigkeit betroffen sein oder eine ungerechte Interaktion beobachten. Mit jeder der drei Perspektiven gehen verschiedene Reaktionen einher, welche unterschiedlich stark ausgeprägt vorliegen können. Personen in der Täterrolle können starke, schwache oder keine Schuldgefühle für einen ungerecht erlangten Vorteil entwickeln, Personen in der Opferrolle können Traurigkeit, Wut und Rache empfinden oder auch relativ unbetroffen bleiben und Personen, welche sich in der Beobachterrolle befinden, können mehr oder weniger gleichgültig reagieren oder aber Empathie für das Opfer fühlen (Schmitt, 1996; Schmitt, Gollwitzer, Maes & Arbach, 2005).

Die interindividuelle Variation von Art und Ausprägung der Ungerechtigkeitssensibilität in den verschiedenen Perspektiven geht nun auch mit Variationen anderer Persönlichkeitseigenschaften einher, wie zum Beispiel mit prosozialem Verhalten. Schmitt et al. (2005) konnten zeigen, dass hohe Ausprägungen von Ungerechtigkeitssensibilität in den Reaktionen als Opfer ne-

gativ mit prosozialen Verhaltensweisen korrelieren. Hohe Ausprägungen aus Sicht der beiden anderen Positionen, als Täter oder Beobachter, weisen hingegen einen positiven Zusammenhang mit prosozialem Verhalten auf und können somit einen Erklärungsansatz für den besonders starken Einsatz von altruistischer Bestrafung bilden – zum Beispiel im Rahmen von ökonomischen Spielen.

1.3 Altruistische Bestrafung und Ökonomie

Zwei charakteristische Merkmale von ökonomischen Modellen bestehen in deren expliziter Formulierung sowie dem sich hieraus ergebenden Vorteil von präzisen hypothetischen Verhaltensvorhersagen. Somit bieten sie eine stabile Grundlage zur Entwicklung von Theorien über menschliche Entscheidungsfindung unter verschiedenen Umständen (Camerer, Loewenstein & Prelec, 2005; Sanfey, Loewenstein, McClure & Cohen, 2006), beispielsweise zur Klärung der Frage, warum Menschen gewillt sind altruistisch zu bestrafen.

1.3.1 Eigennutz in der Ökonomie – Der Mensch als Homo oeconomicus

Ein Grundsatz traditioneller ökonomischer Theorien lautet, dass der Akteur in Entscheidungssituationen jene Handlungsmöglichkeit wählen wird, bei welcher die Maximierung eigener materieller Größen, wie zum Beispiel Geld oder Konsumgüter, am wahrscheinlichsten ist. Zum einen müssen dabei potentielle Handlungsoptionen evaluiert und nach individueller Präferenz selektiert werden (1). Zum anderen gilt es entsprechende Optionen nach ihrer Wahrscheinlichkeit zu bewerten (2), also ob eine qualitativ zunächst am besten erscheinende Wahl auch praktisch umsetzbar ist (Allingham, 2002; Sanfey et al., 2006).

Ein nach ökonomischen Prinzipien und somit streng rational handelnder Mensch, der Homo oeconomicus, würde daher niemals altruistisch handeln und auch nicht gewillt sein altruistisch zu bestrafen. Denn hierbei ist eine offensichtlich negative Konsequenz (1), das Investieren eigener materieller Kosten, sowohl nach der Definition des biologischen als auch des psychologischen Altruismus zu 100 Prozent wahrscheinlich (2).

1.3.2 Fairness in der Ökonomie – Die Theorie der sozialen Präferenzen

Im Rahmen neoklassischer Nutzenmaximierungstheorien kann ein nach ökonomischen Prinzipien geleitetes Verhalten auch in der Fokussierung anderer, nicht materieller Güter oder Ziele bestehen. Unter diesem Gesichtspunkt lassen sich auch soziale Präferenzen von Gleichheit (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) und reziproker Fairness (Falk & Fischbacher, 2001; Levine, 1998; Rabin, 1993) modellieren.

1.3.2.1 Ungleichheitsaversion

Fehr & Schmidt (1999, Seite 819) definieren Fairness als „self-centered inequity aversion“. Diese Abneigung gegenüber Ungleichheit wird von verschiedenen Autoren als ein intrinsisch motivierter Widerstand einer Person gegen die ungleiche Verteilung eines materiellen Guts zwischen sich selbst und anderen (Fehr & Schmidt, 1999) beziehungsweise zwischen sich selbst und dem Gruppendurchschnitt (Bolton & Ockenfels, 2000) betrachtet. Dabei wird sowohl die durch eine Verteilung gegebene Besserstellung als auch eine erlangte Schlechterstellung von der Person als aversiv empfunden. Ausgehend von dem Bedürfnis nach selbstbezogener Fairness ist diese Person somit dazu gewillt einen bestimmten Einsatz eigener materieller Kosten zu leisten. Im Fall der eigenen Schlechterstellung könnte das Ziel einer annähernd gleich gestalteten Aufteilung durch den Einsatz von altruistischer Bestrafung geschehen (siehe auch 1.2.2.2).

Aufgrund des oben erwähnten Aspektes der Eigenbezogenheit, also der situativen Beschränktheit der angestrebten Fairness auf die Verteilung von eigenen materiellen Gütern relativ zu anderen, können Annahmen basierend auf dem Konzept der Ungleichheitsaversion jedoch nicht erklären, warum altruistische Bestrafung zur Herstellung von gleich verteilten Gütern auch durch Dritte (Third Party) geschieht (siehe z.B. Falk & Fischbacher, 2006). Weiterhin finden in diesem Modell die einer Handlung unterliegenden Intentionen eines Interaktionspartners keine Berücksichtigung. Dass diese durchaus wichtige Prädiktoren zur Vorhersage von altruistischer Bestrafung bilden, wird von den Vertretern des folgenden Ansatzes der reziproken Fairness, als eine weitere Theorie zu sozialen Präferenzen, postuliert.

1.3.2.2 Reziproke Fairness

Nach dem Modell der reziproken Fairness wird folgendermaßen argumentiert: „If somebody is being nice to you, fairness dictates that you be nice to him. If somebody is being mean to you, fairness allows – and vindictiveness dictates – that you be mean to him“ (Rabin, 1993, Seite 1281). Wenn faires Verhalten die Referenz dafür bildet, dass ein davon abweichendes Verhalten bestraft wird, bedarf es einer präzisen Klärung dafür, wann ein Verhalten als fair beziehungsweise unfair wahrgenommen wird. Zwei hierfür entscheidende Aspekte bilden nach Falk und Fischbacher (2006) die Konsequenz der Handlung sowie die einer Handlung zu Grunde liegenden Intentionen. Auch wenn „an equitable share of payoffs seems to be the reference standard to determine what is a fair or unfair offer“ (Falk & Fischbacher, 2006, Seite 296), können davon abweichende Angebote durch eine entgegengesetzt gerichtete Intention ausgeglichen werden und noch immer als fair gelten (Fairness = Intention x Konsequenz). Dieser Sachverhalt wurde bereits unter 1.1.3.2 im Zusammenhang mit dem Konzept der starken Reziprozität anhand eines Beispiels illustriert.

1.3.2.3 Die Theorie der sozialen Präferenzen – Zusammenfassung

Die Theorie der sozialen Präferenzen lässt sich als eine Erweiterung der traditionellen Modelle der Ökonomie (siehe 1.3.1) betrachten. Die persönliche Nutzenmaximierung muss hier nicht nach einem Homo oeconomicus und der Maximierung von materiellen Gütern bestehen, sondern zieht auch die Präferenz von fairen Aufteilungen des Akteurs in Betracht. Unter diesem Gesichtspunkt bilden Modelle der sozialen Präferenz eine extern validere Vorhersage vieler Verhaltensweisen. So lässt sich zum Beispiel der Einsatz von altruistischer Bestrafung als eine Antwortreaktion auf das Zusammenspiel von ungleicher Aufteilung und egoistischer Intention des Interaktionspartners verstehen (Falk & Fischbacher, 2006). Weiterhin bildet die Theorie der sozialen Präferenzen einen fließenden Übergang von ultimaten und proximaten Erklärungsmodellen für altruistisches Verhalten. Ultimate beziehungsweise evolutionäre Grundsätze von eigener Nutzenmaximierung sowie proximate Modelle von sozialer Norm und kognitiver Bewertung finden schlüssige Integration. Emotionen oder Empathie werden zwar weder im Konzept der Ungleichheitsaversion noch nach dem Prinzip der reziproken Fairness berücksichtigt, doch können diese als mediierende Variablen des Zusammenhangs von Fairnessbruch und Ausführung von altruistischer Bestrafung betrachtet werden (siehe 1.2.2.2 und 1.2.2.3).

Das erweiterte Nutzenprinzip der Theorie der sozialen Präferenzen basiert unter anderem auf den Ergebnissen zahlreicher Studien, welche mittels spieltheoretischer Analysen versuchen menschliches Entscheidungsverhalten näher zu untersuchen. Im Folgenden wird das Prinzip der Spieltheorie kurz vorgestellt sowie eines der vielen verfügbaren Paradigmen näher betrachtet (zur Übersicht siehe z.B. Camerer, 2003), das *Dictator Game Paradigma*.

1.3.3 Das Prinzip der Spieltheorie

Die Spieltheorie versucht Entscheidungen mehrerer Akteure in sozialen Konfliktsituationen, wie der gemeinsamen Auf- oder Zuteilung eines materiellen Guts, durch mathematische Modelle abzubilden und für präzise Verhaltensvorhersagen zu nutzen. In der Regel handelt es sich dabei um interdependente Entscheidungssituationen von rationalen Spielern mit festgelegten Präferenzen. Die Entscheidung des Spielers A geschieht unter Bezugnahme der Präferenzen des Spielers B; Spieler A handelt strategisch. Die theoretische Vorhersage nach ökonomischen Modellen basiert auf dem Prinzip des Homo oeconomicus (siehe 1.3.1) und beinhaltet die Maximierung eigener materieller Güter (Holler, 2003).

1.3.4 Paradigmen der Spieltheorie

Die als Paradigmen bezeichneten Spiele der Spieltheorie beinhalten Situationen, unter welchen die Akteure Entscheidungen treffen. Je nach Untersuchungsgegenstand kann ein Paradigma aus zwei, drei oder mehreren Spielern bestehen, die einmalige (*One Shot*) oder mehrfache (*Repeated*) Interaktion (unter denselben oder immer neuen Spielern) vorsehen und die einzelnen Spieler aus verschiedenen Perspektiven (First Person, Third Party) agieren lassen. Merkmale, welche typischerweise alle Spiele gemeinsam haben, sind die wechselseitige Anonymität der Spieler, der Ausschluss von Kommunikation unter den Spielern sowie die finanzielle Entschädigung der Teilnahme in Abhängigkeit vom Spielergebnis (Camerer, 2003; Holler, 2003).

Im Folgenden wird das in der eigenen Studie verwendete Dictator Game Paradigma, zum einen in seiner klassischen Form und zum anderen in einer erweiterten Form mit Bestrafungsoption, im Detail betrachtet.

1.3.4.1 Das Dictator Game

Das klassische Dictator Game modelliert eine Situation, in welcher der *Dictator* eine Entscheidung darüber zu treffen hat, wie viel dem *Empfänger* von einem bestimmten Ausgangsbetrag zugeteilt werden soll. Nach der Annahme der traditionellen Spieltheorie gibt ein Dictator nichts von dem aufzuteilenden Betrag an den Empfänger, denn nur so kann er seinen eigenen materiellen Nutzen maximieren. Die Ergebnisse empirischer Studien zeigen jedoch, dass nur in circa 20 Prozent der Fälle eine Null-Zuteilung getroffen wird; in weiteren 20 Prozent wird hingegen sogar der halbe Betrag geteilt (Forsythe, Horowitz, Savin & Sefton, 1994; Hoffman, McCabe & Smith, 1996). Nach einer Metaanalyse von Camerer (2003) besteht die durchschnittliche Zuteilung des Dictators an den Empfänger in etwa einem Fünftel des Ausgangsbetrages.

In der Regel wird das Dictator Game als One Shot-Game gespielt, bei welchem der Dictator aktiv agiert, der Empfänger hingegen vollkommen passiv ist und keinen Widerspruch leisten kann. Besonders im Vergleich mit dem Ultimatum Game, welches sich vom Dictator Game darin unterscheidet, dass eine Zuteilung vom Empfänger abgelehnt werden kann, wird die Höhe der Zuteilung im Dictator Game oft als Maß für Altruismus betrachtet (zum Vergleich von Dictator Game und Ultimatum Game siehe z.B. Camerer, 2003; Forsythe et al., 1994).

1.3.4.2 Das Dictator Game mit Bestrafungsoption

In einer erweiterten Variante des Dictator Games besteht die Möglichkeit eine durch den Dictator getroffene Zuteilung zu bestrafen. Dies kann entweder durch den Empfänger direkt (First Person) oder indirekt durch einen Beobachter geschehen (Third Party). Die Bestrafung wird anhand eines bestimmten Punktebetrages ermöglicht, dessen Einsatz sowohl für den bestrafenden Spieler als auch für den bestraften Dictator mit materiellem Verlust einhergeht. Ausgehend von den Annahmen der traditionellen Spieltheorie wird weder aus der First Person noch aus der Third Party Perspektive bestraft, weil die Bestrafung kostspielig ist und somit entgegen der eigenen Nutzenmaximierung wirkt. Doch auch in diesem Fall sprechen die empirischen Ergebnisse gegen die theoretische Vorhersage. Die meisten Spieler bestrafen die Dictator für Zuteilungen, welche kleiner als der halbe Ausgangsbetrag sind, und zwar aus der First Person und Third Party Perspektive vergleichsweise hoch (Fehr & Fischbacher, 2004b; Strobel, Zimmermann, Schmitz, Reuter, Gallhofer, Windmann & Kirsch, in prep.).

Doch Verhaltensweisen, welche sich auf behavioraler Ebene gleichen, können unterschiedliche neuronale Prozesse zu Grunde liegen. Verschiedene Vorgänge im Gehirn sind wiederum häufig mit jeweils anderen Motiven für Verhalten assoziiert. Die Untersuchung der neuronalen Korrelate von altruistischer Bestrafung aus verschiedenen Perspektiven kann somit bedeutsam zur Motivspezifikation beitragen. Unter Verwendung von *funktioneller Magnetresonanztomographie* (siehe 1.4.3) konnten Strobel et al. (in prep.) zeigen, dass bestimmte Regionen des Gehirns eine höhere Aktivität aufweisen, wenn aus der First Person Perspektive agiert wurde relativ zur Third Party Perspektive (anteriorer und posteriorer cingulärer Cortex, Nucleus accumbens) und wenn bestraft wurde relativ zu keiner Bestrafung (Nucleus accumbens, Nucleus caudatus, Insula, dorsolateraler präfrontaler Cortex, anteriorer cinulärer Cortex).

Mögliche Funktionen dieser Regionen des Gehirns im Kontext von altruistischer Bestrafung werden unter 1.4.4 näher erläutert. Zunächst soll ein Bereich der Forschung vorgestellt werden, in welchem genau diese Integration des Einsatzes von bildgebenden Methoden der Neurowissenschaft und Paradigmen der ökonomischen Spieltheorie stattfindet. Dieser Bereich wird als Neuroökonomie bezeichnet und bildet den wissenschaftlichen Rahmen der eigenen Studie.

1.4 Altruistische Bestrafung, Ökonomie und Neuroforschung

1.4.1 Das neue Forschungsfeld der Neuroökonomie

Hinter dem Konzept der Neuroökonomie verbirgt sich ein noch recht junger, integrativer Ansatz von Grundsätzen der Psychologie, der Neurowissenschaft und der Ökonomie mit dem Ziel, neue Erklärungsmodelle in Bezug auf menschliches Entscheidungs- und Wahlverhalten zu spezifizieren. Das neue Forschungsfeld wurde anfänglich sowohl von den Vertretern der neurowissenschaftlichen Methodik als auch von Seiten der Ökonomie mit großer Unsicherheit bezüglich eines gemeinschaftlichen Erkenntnisgewinnes betrachtet. So zeigten sich Ökonomen seit jeher als „skeptical of the ability of ‚process measures' to contribute to our understanding of economic and social behavior [und Neurowissenschaftler bewerteten] economics as too abstract and removed from the mechanisms of interest in the brain" (Sanfey et al., 2006, Seite 108). Doch schon heute, circa zwei Jahrzehnte nach den ersten Berührungspunkten beider Disziplinen, zeigt sich der Wert jener wissenschaftlichen Symbiose. So können zum Beispiel klar definierte, mathematische Modelle aus der Ökonomie als Richtlinien für standardisierte *Verhaltens-Baselines*

in der neurowissenschaftlichen Forschung gelten und somit bei der Interpretation von neurobiologischen Befunden helfen. Weichen empirische Verhaltensweisen auf behavioraler Ebene systematisch von jenen theoretischen Vorhersagen ab, können wiederum mit Blick auf neuronale Verhaltenskorrelate Hinweise hierfür gefunden und ursprüngliche Modelle angepasst oder gegebenenfalls neu formuliert werden (Camerer et al., 2005; Sanfey et al., 2006).

In diesem Zusammenhang lässt sich auch das Phänomen der altruistischen Bestrafung im Licht eines multiplen Erklärungsansatzes neu betrachten.

1.4.2 Multiple Ansätze zur Entscheidungsfindung

Einen bedeutsamen Beitrag, welchen die Psychologie für die Neuroökonomie leistet, bildet die Unterscheidung zwischen automatischen und kontrollierenden Prozessen zur Verhaltenssteuerung (Posner & Snyder, 1975; Schneider & Shiffrin, 1977). Automatische Prozesse laufen schnell und effizient ab, können daher oft parallel erfolgen, doch sind in der Regel auf spezifische Handlungsvorgänge beschränkt und somit relativ unflexibel. Es wird vermutet, dass diese Prozesse überlernte beziehungsweise internalisierte Verhaltensweisen initiieren und oft unbewusst passieren. Dem gegenüber stehen kontrollierende Prozesse, für welche es der Beanspruchung höherer kognitiver Mechanismen bedarf. Kontrollprozesse laufen daher in der Regel etwas langsamer als automatische Prozesse ab, sind dafür jedoch hochflexibel und in der Lage, bewusst zur Steuerung von Verhalten in verschiedensten Situationen beizutragen (Sanfey et al., 2006). Die Differenzierung in automatische und kontrollierende Prozesse, welche mehr als die zwei Enden eines Kontinuums von neuronalen Prozessen verstanden werden können als eine reine Dichotomie (Kahneman & Treisman, 1984), hat bereits Einfluss auf die Untersuchung von menschlichem Entscheidungsverhalten gezeigt. Bezüglich der Wahl einer unter mindestens einer weiteren alternativen Handlung bedarf es deren Bewertung hinsichtlich eigener Prämissen. Hierfür werden zwei neuronale Systeme unterschieden. System 1 ähnelt den Charakteristika von automatisch ablaufenden Prozessen. Es impliziert eine schnelle neuronale Antwort, meist auf der Basis von Heuristiken ablaufend, und scheint für intuitive Antworten verantwortlich zu sein. System 2, welches stark mit den kontrollierenden Prozessen assoziiert ist, evaluiert die Antwort durch System 1 und korrigiert oder überschreibt diese gegebenenfalls (siehe z.B. Kahneman, 2003).

System 2 erinnert in seiner Kontroll- und Evaluationsfunktion an das stark rationalistische Prinzip eines Homo oeconomicus mit seinen klaren, materialistischen Präferenzen in der klassischen Spieltheorie. Wie bereits an verschiedenen Stellen erwähnt, können ökonomische Modelle das menschliche Verhalten in Entscheidungssituationen nur zum Teil erklären. Ebenso bilden auch die Kontrollprozesse ausgehend von System 2 nur einen Bestandteil aller neuronalen Prozesse und können je nach Situation schnelle automatische Prozesse von System 1 noch „auffangen“ oder werden von diesen „überwältigt“ (Sanfey et al., 2006).

Im Rahmen der neuroökonomischen Forschung haben Ökonomen bereits begonnen die Grenzen traditioneller, strikt rationaler Modelle dahingehend zu erweichen, dass sie neben kontrollierenden auch automatische Prozesse in neue Modelle integrieren. Benhabib und Bisin (unveröffentlicht, zitiert nach Sanfey et al., 2006) postulieren ein Modell zur Erklärung von Entscheidungsfindung, nach welchem automatische Prozesse anfängliches Verhalten determinieren. Das Dominieren jener schnellen, leichter verfügbaren Prozesse wird jedoch nur so lange zugelassen, bis möglicherweise hierdurch entstehende Kosten ein bestimmtes Niveau übersteigen und den Einsatz ressourcenintensiverer Kontrollmechanismen rechtfertigt. Dieser Aspekt der Kostenlimitierung wird auch von Fehr & Fischbacher (2003) in Bezug auf mögliche Grenzen von altruistischem Verhalten aufgriffen. Der Vergleich mag auf den ersten Blick recht abstrakt erscheinen. Doch wird altruistisches Verhalten als Resultat verschiedener, zunächst dominierender automatischer Prozesse betrachtet und die Begrenzung des altruistischen Einsatzes als Verhaltenskorrelat zunehmender neuronaler Kontrollmechanismen interpretiert, wird der Zusammenhang plausibel.

Mit Hilfe von bildgebenden Methoden der Neurowissenschaft ist es seit einiger Zeit möglich, den relativen Einfluss beider neuronaler Systeme auf menschliche Entscheidungsfindung näher zu untersuchen. Dabei interessieren insbesondere solche Situationen, in welchen Kontrollprozesse und emotionale Prozesse, als eine Kategorie von automatischen Prozessen, um die dominierende Antwort konkurrieren. Solch eine Situation könnte auch der Einsatz von altruistischer Bestrafung im Dictator Game darstellen und wird in der vorliegenden Studie unter Verwendung von einer bildgebenden Methode der Neurowissenschaft, der funktionellen Magnetresonanztomographie, näher betrachtet.

1.4.3 Die Methode der funktionellen Magnetresonanztomographie

Die Magnetresonanztomographie (MRT) wurde vor knapp dreißig Jahren zunächst als bedeutsames anatomisches Diagnoseinstrument in den klinischen Kontext eingeführt und leistet seit kürzerer Zeit unter Anwendung von funktionellen Messungen (fMRT) einen entscheidenden Beitrag für die neurowissenschaftliche Forschung. Das 1990 von Ogawa und Kollegen entdeckte Prinzip der fMRT (Ogawa, Lee, Kay & Tank, 1990) basiert auf der Messung eines Signals, welches vom Sauerstoffgehalt im Blut abhängt (*Blood Oxygen Level Dependency*, BOLD). Es konnte gezeigt werden, dass die Zu- und Abnahme der Blutoxygenierung eng an Veränderungen der lokalen neuronalen Aktivität entsprechender Hirnbereiche gekoppelt ist (Logothetis, Pauls, Augath, Trinath & Oeltermann, 2001). In Kombination mit ausgewählten Paradigmen zur Operationalisierung von jeweils interessierenden Verhaltens- oder Reizphänomenen lassen sich somit Erkenntnisse über die Lokalisation und Dynamik von Hirnprozessen gewinnen. Im Gegensatz zu anderen bildgebenden Methoden wie der PET oder der *Computertomographie* bedarf es weder für die anatomischen noch für die funktionellen Aufnahmen des Einsatzes von radioaktiven Strahlen oder der Verabreichung von Tracer-Substanzen. Die Methode der fMRT gilt daher als nicht invasiv und eignet sich hervorragend für Forschungszwecke im Humanbereich (für einen ausführlichen Überblick zur Methode und zu physikalischen Grundlagen der fMRT siehe z.B. Buxton, 2002).

1.4.4 Neuronale Korrelate von altruistischer Bestrafung

Mit Gründung des Forschungsbereichs der Neuroökonomie (siehe 1.4.1) wurde ebenso der Grundstein verschiedener neuer Modelle zur Erklärung von menschlicher Entscheidungsfindung gelegt. Mit Bezug auf aktuelle Erkenntnisse der Neurowissenschaft konnten in der psychologischen Forschung funktional differenzierbare neuronale Bereiche spezifiziert werden (siehe 1.4.2). Diese lassen sich wiederum mit Hilfe bildgebender Methoden visualisieren (siehe 1.4.3).

Spätestens mit Blick auf die Dynamik von Hirnprozessen wird klar, dass sich die menschliche Entscheidungsfindung im sozialen Kontext nicht auf einzelne Regionen des einen oder anderen neuronalen Systems zurückführen lässt. Weiterhin sind die neuronalen Aktivierungen bestimmter Hirnregionen niemals separat und absolut, sondern immer relativ zu den Aktivierungen anderer Hirnregionen und im Vergleich verschiedener Bedingungen zu betrachten. Diese Anmerkungen gilt es zu berücksichtigen, auch wenn im Folgenden eine grobe funktionale Klassifizierung ein-

zelner Hirnregionen vorgenommen wird. Es wird ein Überblick solcher Regionen des Gehirns gegeben, welche, mit Verweis auf die Ergebnisse verschiedener bildgebender Studien, in ihrer Interaktion eine bedeutsame Rolle für die Erklärung von altruistischer Bestrafung zu spielen scheinen. Diese Regionen sind vor allem die Insula und das Striatum, Bereiche des frontalen Neocortex sowie die cingulären Cortices.

1.4.4.1 Die Insula

Die Insula besteht aus drei Zonen und liegt im Gehirn tief in der Fissura lateralis, verdeckt von den frontalen und parietalen Opercula sowie dem Lobus temporalis. In verschiedenen fMRT Studien konnte die Aktivierung der Insula in Zusammenhang mit Schmerz und Distress (Derbyshire, Jones & Gyulai, 1997; Evans, Banzett, McKay, Frackowiak & Corfield, 2002; Iadarola, Berman, Zeffiro, Byas-Smith, Gracely, Max & Bennett, 1998), Hunger und Durst (Denton, Shade, Zamarippa, Egan, Blair-West, McKinley, Lancaster & Fox, 1999; Tataranni, Gautier, Chen, Uecker, Bandy, Salbe, Pratley, Lawson, Reiman & Ravussin, 1999) und autonomer Erregung (Critchley, Elliott, Mathias & Dolan, 2000) gebracht werden. Es wurde weiterhin auf die Bedeutung dieser Hirnregion in Bezug auf Empathie und Emotionen verwiesen (Vignemont & Singer, 2006; de Waal, 2007). Hierbei scheint die anteriore Insula bei der Repräsentation und Evaluation von spezifischen negativen Emotionen eine Rolle zu spielen (Calder, Lawrence & Young, 2001). Zwei dieser Emotionen sind Ärger und Ekel, wobei letztere im Kontext von sozialem Normbruch eher als *sozialer Ekel* zu interpretieren ist (Damasio, Grabowski, Bechara & Damasio, 2000; Phillips, Young, Senior & Brammer, 1997; Sanfey, Rilling, Aronson, Nystrom & Cohen, 2003). Sanfey et al. (2003) konnten in einer fMRT Studie demonstrieren, dass die Aktivierung in der Insula mit der Tendenz einer Person zu altruistischer Bestrafung im Ultimatum Game positiv korreliert. Solche Probanden lehnten unfaire Angebote in höherem Maße ab und bestraften somit stärker altruistisch, welche ebenso eine stärkere Aktivierung der Insula aufwiesen. Dieses Ergebnis bietet Hinweise für die Annahme, dass eine Entscheidung für altruistische Bestrafung durch negative Emotionen oder die Wahrnehmung von sozialem Ekel mediiert werden kann (siehe 1.2.2.2, Emotionen negativer Valenz).

Der Einsatz von altruistischer Bestrafung könnte aber auch durch die antizipierte Freude darüber, „gerechtfertigt“ bestrafen zu dürfen beziehungsweise es einer unfair gehandelten Person „heimzahlen“ zu können (siehe 1.2.2.2, Emotionen positiver Valenz), motiviert sein. Dieses

Motiv müsste sich in einer Aktivierung von Regionen des neuronalen Belohnungssystems widerspiegeln, so zum Beispiel im Striatum.

1.4.4.2 Das Striatum

Der Nucleus caudatus (N. caudatus), das Putamen und der Nucleus accumbens (N. accumbens) bilden gemeinsam das Striatum und somit die größte subkortikale Zellmasse im menschlichen Telencephalon. Der Striatum-Komplex lässt sich in einen dorsalen und einen ventralen Teil untergliedern. Das dorsale Striatum setzt sich zusammen aus den dorsalen und dorsolateralen Teilen des N. caudatus und des Putamens; das ventrale Striatum umfasst den gesamten N. accumbens sowie benachbarte ventrale und mediale Teile des N. caudatus und des Putamens (Nieuwenhuys, Voogd & Huijzen, 1991). Bezüglich der Bedeutung von Belohnung auf die menschliche Entscheidungsfindung heben einige Autoren eher den dorsalen Teil (z.B. Balleine, Delgado & Hikosaka, 2007; de Quervain et al., 2004) und andere Autoren eher den ventralen Teil (z.B. Harbaugh, Mayr & Burghart, 2007; Kable & Glimcher, 2007; O'Doherty, 2004) des Striatums hervor. Einigkeit herrscht jedoch darüber, dass das Striatum einen wichtigen Bestandteil eines neuronalen Belohnungsnetzwerkes bildet. Die Ergebnisse verschiedener bildgebender Studien geben Hinweise darauf, dass eine Aktivierung in dieser Hirnregion unter anderem mit einer Bewertung der subjektiven Valenz einer Belohnung einhergeht (z.B. Kable & Glimcher, 2007). Die Bedeutung des Striatums oder Teilen des Striatums im Entscheidungsprozess besteht weiterhin darin, zielgerichtet solche Handlungen zu wählen, welche durch eine antizipierte Belohnung motiviert sind (z.B. Balleine et al., 2007; O'Doherty, 2004). In Bezug auf das Motiv der Freude und Belohnung für altruistische Bestrafung gehen de Quervain et al. (2004) sogar soweit anzunehmen, „that the caudate plays a decisive role in altruistic punishment". Sie konnten zeigen, dass solche Personen mit der stärksten N. caudatus Aktivierung bei kostenfreier Bestrafung auch diejenigen Personen waren, welche am meisten bestraften, wenn die Bestrafung kostenintensiv wurde. Daraus zogen sie den Schluss, dass „high caudate activation seems to be responsible for a high willingness to punish, which suggests that caudate activation reflects the anticipated satisfaction from punishing defectors" (de Quervain et al., 2004, Seite 1258).

Neben dem Striatum spielen weiterhin auch der posteriore cinguläre Cortex (PC) sowie die orbitalen als auch medialen Regionen des präfrontalen Cortex eine Rolle im neuronalen Belohnungssystem (Kable & Glimcher, 2007; Rushworth, 2008).

1.4.4.3 Der frontale Neocortex

Nach der topographischen Klassifizierung von Petrides & Pandya (2003) setzt sich der frontale Cortex aus dem dorsolateralen und ventrolateralen präfrontalen Cortex, dem orbitalen frontalen, frontopolaren und dem medialen frontalen Cortex zusammen. Jeder dieser Cortices besteht erneut aus mehreren Arealen. Anhand von Läsionsstudien konnte gezeigt werden, dass insbesondere der präfrontale Cortex (PFC) bei der antizipatorischen Einstellung, Planung und Initiative von Handlungen beteiligt ist (Milner, Petrides & Smith, 1985).

Im Zusammenhang mit sozialen Normen scheint die Region des PFC zum einen bei der Repräsentation entsprechender Normen als auch bei der Entscheidung für die Bestrafung von normwidrigen Verhalten eine wichtige Rolle zu spielen. Knoch, Pascual-Leone, Meyer, Treyer & Fehr (2006) konnten zeigen, dass eine durch transkranielle Magnetstimulation (TMS, zur Methode der TMS siehe z.B. Walter, 2005, Kapitel 11) herbeigeführte leichte Herabsetzung der Aktivität im rechten PFC dazu führte, dass die Probanden signifikant weniger oft ein vom Spielpartner gemachtes unfaires Angebot ablehnten sowie ihre Entscheidung hierfür schneller trafen. Da die gehemmte neuronale Aktivierung im PFC nicht zu einer Veränderung des Fairnessurteils an sich führte, kann weiterhin angenommen werden, dass die Fairness-Evaluation einer gegebenen sozialen Interaktion vorwiegend von anderen Regionen ausgeht. Diese Ergebnisse sprechen für eine kognitive, zeitaufwändige Kontrollfunktion des PFC zur Umsetzung einer Handlung und zwar eher nach dem Prinzip der reziproken Fairness (siehe 1.3.2.2), als nach dem Prinzip des Homo oeconomicus (siehe 1.3.1). Die Autoren anderer neuroökonomischer Studien verweisen zudem auf die Bedeutung der dorsolateralen Region des präfrontalen Cortex (DLPFC). Insbesondere der rechte (r)DLPFC scheint sowohl die Entscheidung für oder gegen den Einsatz von Bestrafung bei kostenintensiver Bestrafungsoption aus der First Person Perspektive (z.B. Sanfey et al., 2003) als auch bei kostenfreier Bestrafung aus der Third Party Perspektive (z.B. Buckholtz, Asplund, Dux, Zald, Gore, Jones & Marois, 2008) mit zu moderieren. Eine relativ erhöhte Aktivierung in dieser Region kann somit zunächst als unabhängig von der Handlungsperspektive gesehen werden. Dieser Befund, wie auch die Tatsache, dass der rDLPFC auch bei kostenfreier Bestrafung von unfairem Verhalten erhöht aktiviert ist, spricht erneut eher für die Kontrollfunktion zur Umsetzung von reziproker Fairness, als zur Fokussierung von materiellem Eigennutz. Letztere Bedeutung des DLPFC im Zusammenhang mit altruistischer Bestrafung vermuten jedoch Sanfey et al. (2003) nach den Ergebnissen ihrer bereits unter 1.4.4.2 aufgeführten Studie.

Eine relativ zur Insula erhöhte Aktivierung des DLPFC führte dazu, dass die Probanden auch unfaire Angebote im Ultimatum Game annahmen. Dies veranlasste dazu, die Aktivierung des DLPFC als „reflecting the steady task representations of money maximization“ zu interpretieren (Sanfey et al., 2003, Seite 1757).

Einigkeit bezüglich einer Funktion des PFC beziehungsweise des (r)DLPFC herrscht jedoch darüber, dass diese Region des Gehirns eine Art Kontrollinstanz für die Umsetzung einer potentiellen Handlung bildet. Je nach Argumentationsstrang sollen dabei zum einen „fundamental impulses associated with self-interest to be controlled in order to maintain and to implement culture-dependent fairness goals“ (Knoch et al., 2006, Seite 829) und zum anderen „emotional areas in influencing the decision [kontrolliert werden mit dem Ziel] of accumulating as much money as possible“ (Sanfey et al., 2003, Seite 1757).

Zu Beginn des Abschnitts wurde bereits darauf verwiesen, dass eine klare funktionale Zuteilung von Hirnregionen zu dem einen oder dem anderen System stets mit Vorsicht zu betrachten ist. Basierend auf neuroökonomischen Befunden kann dennoch eine grobe Klassifizierung der aufgeführten Hirnregionen vorgenommen werden. So können die mit emotionsinduzierten Handlungen verknüpften Regionen der Insula und des Striatums eher dem automatischen System 1 zugeordnet werden. Die Regionen des PFC, welche stärker exekutive Funktionen zur Handlungssteuerung innehaben zu scheinen, werden eher als Vertreter des kontrollierenden Systems 2 gesehen. Konkurrieren kognitiv kontrollierende und emotionale Motive um die dominante Antwort, so spiegelt sich dieser neuronale Konflikt in anderen Regionen des Gehirns wider, zum Beispiel im anterioren cingulären Cortex.

1.4.4.4 Der anteriore cinguläre Cortex

Der anteriore cinguläre Cortex (ACC) wird anatomisch zum medialen frontalen Cortex gezählt (Rushworth, 2008). Bezüglich seiner Funktion wird ihm eine wichtige Rolle bei der Erkennung von Fehlern, welche mit einer aktuellen oder potentiell angestrebten Handlung einhergehen können, zugesprochen. Fehlerhaftes Handeln wiederum geht meist mit eigenen Kosten, sei es in psychologischer oder materieller Form, einher. Sanfey et al. (2003) verweisen auf einen Anstieg der ACC Aktivierung bei der Konfrontation mit unfairen Angeboten relativ zu fairen Angeboten im Ultimatum Game. Sie interpretieren diesen Aktivierungsunterschied als Resultat eines neuronalen „conflict between cognitive and emotional motivations in the Ultimatum Game“ (Sanfey et al., 2006, Seite 1757). In Unterabschnitt 1.4.4.3 wurde bereits dargestellt, dass die ko-

gnitive Motivation darin bestehen könnte, ein Handeln zum einen nach den Prinzipien der Fairnessnorm oder zum anderen nach den Prinzipien des Homo oeconomicus umzusetzen. Droht die kognitive Kontrolle durch emotionale Motivationen gefährdet zu werden, können im ersten Fall psychologische und im zweiten Fall materielle Kosten die Folgen sein. Eine emotional geleitete Handlung kann jedoch auch Nutzen bringen, sei es durch den Abbau von Spannung und Ärger (siehe 1.4.4.1) oder die Wahrnehmung von Freude und Erleichterung (siehe 1.4.4.2). Der ACC ist mit dafür verantwortlich, die von verschiedenen Systemen ausgehenden neuronalen Antworten zu integrieren, nach deren Valenz und individueller Präferenz zu gewichten, um somit ein Handeln nach dem Prinzip der relativen Nutzenmaximierung gewährleisten zu können (Rushworth, 2008; Sanfey et al., 2003).

Bestrafung von unfairem Verhalten in Paradigmen der Spieltheorie wird auch dann gezeigt, wenn diese Bestrafung mit materiellen Kosten verbunden ist. Daher ist anzunehmen, dass zumindest ein großer Teil des relativen Nutzens bei einer Entscheidung für diese Handlung in der Maximierung von sozialen oder intrinsischen Präferenzen und nicht in der Maximierung von materiellen Präferenzen besteht. Das Ziel der vorliegenden Studie ist es, die intrinsischen Motive für die Entscheidung zu altruistischer Bestrafung unter Bezug von behavioralen, neuronalen sowie differentialpsychologischen Aspekten näher zu spezifizieren. Neben eines vermutetes Rachemotivs in der First Person Bedingung soll insbesondere die Bedeutung eines Fairnessmotives sowohl in der First Person als auch in der Third Party Bedingung systematisch untersucht werden.

1.5 Altruistische Bestrafung in der vorliegenden Studie

Im Dictator Game der vorliegenden Studie befinden sich die Probanden je nach Bedingung in der Rolle des Empfängers oder in der Rolle eines Beobachters einer Zuteilung durch den Dictator (siehe 1.3.4.1). Sie haben dabei sowohl aus der First Person Perspektive als auch aus der Third Party Perspektive in der Hälfte aller Durchgänge die Möglichkeit den Dictator unter Einsatz von Strafpunkten zu bestrafen (siehe 1.3.4.2). Während der übrigen Durchgänge sollen sie angeben, als wie fair oder unfair sie eine entsprechende Zuteilung einschätzen. Dieses Fairnessurteil ist kostenfrei. Normkonformes beziehungsweise normverletzendes Verhalten kann im Dictator Game anhand der Höhe der Zuteilungen durch den Dictator operationalisiert werden. Als Referenz für eine faire Aufteilung gilt der *Equal Split*, da weder die Akteure in der Rolle der Dictator

noch die Akteure in der Rolle der Empfänger oder Beobachter sich einen erhöhten materiellen Output durch zuvor geleisteten materiellen Input gerechtfertigt haben (siehe 1.2.2.1). Eine ausführliche Darstellung des Dictator Games mit Bestrafungsoption und Fairnessurteil erfolgt im Unterabschnitt 2.2.2 des Methodenteils.

1.5.1 Fragestellung und Hypothesen

1.5.1.1 Verhaltensebene

Die Fairnessnorm impliziert objektiv gültige Grundsätze (siehe oben). Diese werden in der vorliegenden Studie als richtungweisend für oder gegen den Einsatz von altruistischer Bestrafung durch den Akteur angesehen. Daher sollte sowohl die Beurteilung der Fairness als auch die Höhe der Bestrafung einer durch den Dictator gemachten Zuteilung unabhängig von der Wahrnehmungsperspektive geschehen und einen hohen positiven Zusammenhang in First Person und Third Party Bedingung aufweisen.

Die Entscheidung für eine Bestrafung, die mit eigenen Kosten verbunden ist, kann im Gegensatz zur Entscheidung nicht zu bestrafen ein neuronales Wechselspiel von zum Teil konkurrierenden Systemen auslösen. Es wird daher angenommen, dass bei Bestrafung längere Reaktionszeiten als bei keiner Bestrafung unabhängig von der Wahrnehmungsperspektive benötigt werden. Des Weiteren sollte sich das zusätzlich zum Fairnessmotiv vermutete Rachemotiv für die Bestrafung in der First Person Bedingung in kürzeren Reaktionszeiten relativ zur Third Party Bedingung äußern.

1.5.1.2 Neuronale Ebene

Auf neuronaler Ebene sollen die Befunde von Strobel et al. (in prep.) repliziert werden. Es werden daher zum einen erhöhte neuronale Aktivierungen in den Regionen des ACCs, des DLPFC, der Insula, des N. caudatus und des N. accumbens erwartet, wenn bestraft wurde relativ zu den Aktivierungen, wenn nicht bestraft wurde. Zum anderen werden in der First Person Bedingung erhöhte Aktivierungen in den Regionen des ACCs, des PCs und des N. accumbens relativ zur Third Party Bedingung angenommen.

Es soll weiterhin untersucht werden, welche neuronalen Korrelate der Bestrafung zu Grunde liegen, wenn die Bestrafungsoption um die Fairnessbeurteilung einer Zuteilung „bereinigt“

wurde. Hierfür wird die erhöhte neuronale Aktivierung unter Bestrafungsoptionen relativ zu der neuronalen Aktivierung unter Fairnessbeurteilungen betrachtet.

1.5.1.3 Differentialpsychologische Aspekte

Neben den allgemeinpsychologischen Fragestellungen ist es ebenfalls im Interesse der vorliegenden Studie eventuell auftretende interindividuelle Unterschiede auf behavioraler oder neuronaler Ebene anhand von unterschiedlichen Ausprägungen in den Persönlichkeitsmerkmalen der Ungerechtigkeitssensibilität, Empathie und Reziprozität erklären zu können.

Zum Ersten wird angenommen, dass ein positiver Zusammenhang zwischen der dispositionellen Ausprägung von Ungerechtigkeitssensibilität in der Rolle des Beobachters und der Höhe der Bestrafung in der Third Party Bedingung besteht (siehe Schmitt et al., 2005). Wir vermuten weiterhin, dass Personen, welche angeben eher sensibel für Ungerechtigkeit zu sein, wenn sie selbst davon betroffen sind, bei der Konfrontation mit unfairen Angeboten eine erhöhte Aktivität der Insula in der First Person relativ zur Third Party Bedingung aufweisen. Personen, welche angeben eher sensibel für Ungerechtigkeit zu sein, wenn sie diese beobachten, sollten bei der Konfrontation mit unfairen Angeboten hingegen eine erhöhte Aktivität der Insula in der Third Party relativ zur First Person Bedingung zeigen.

Zum Zweiten wird ein positiver Zusammenhang zwischen der dispositionellen Ausprägung von Empathieempfinden und der Stärke der neuronalen Aktivierungen sowohl in der Insula als auch im ACC in der Third Party Bedingung erwartet (siehe Hein & Singer, 2008). Wir wollen außerdem untersuchen, ob stark empathische Personen in der Third Party Bedingung auch verstärkt altruistisch bestrafen.

Zum Dritten wird die Annahme eines positiven Zusammenhangs zwischen der dispositionellen Ausprägung von negativer Reziprozität und der Höhe von Bestrafung in der First Person Bedingung getroffen (siehe Perugini, Gallucci, Presaghi & Ercolani, 2003). Personen, die bevorzugt mit „Rache“ oder „Vergeltung“ auf unfaires Verhalten reagieren, sollten zudem eine erhöhte Aktivität in Belohnungsarealen aufweisen, wenn sie in der First Person Bedingung bestrafen.

2 Methode

2.1 Stichprobe

Insgesamt wurden 29 freiwillige Teilnehmer (21 Frauen und 8 Männer) von uns rekrutiert. Davon brachen drei weibliche Probanden die fMRT Messung aufgrund von Platzangst vorzeitig ab. Der Kopf eines männlichen Probanden konnte wegen einer überdurchschnittlichen Schädelgröße nicht adäquat gelagert werden, so dass die Messung nicht durchgeführt werden konnte. Die Datensätze dieser vier Probanden wurden daher a priori von der Analyse ausgeschlossen. Eine weitere Probandin vergab weder in der First Person noch in der Third Party Bedingung Strafpunkte. Da sie somit jenes Phänomen, welches wir in der aktuellen Studie näher betrachten wollen, nicht zeigte und in der statistischen Analyse keine Varianzzerlegung möglich war, ging auch dieser Datensatz nicht in die weitere Analyse ein. Das durchschnittliche Alter der verbleibenden 24 Teilnehmer (19 Frauen und 7 Männer) betrug M = 22.30 Jahre (SE = 1.08). Alle Probanden hatten schon vor dem eigentlichen Messtermin eine von der Abteilung Allgemeine Psychologie II zusammengestellte Fragebogenbatterie ausgefüllt. Diese enthielt eine Auswahl von Fragebogenskalen zur Messung von Persönlichkeitskonstrukten, von welchen angenommen wurde, dass sie einen Einfluss auf das Verhalten im Experiment haben könnten. Des Weiteren wurden potentielle Studienteilnehmer durch ein beigefügtes MRT-Sicherheitsprotokoll bereits über diese Methode informiert sowie über deren Ausschlusskriterien aufgeklärt.

Personen, die angegeben hatten, weder Risikofaktoren aufzuweisen noch an psychiatrischen oder neurologischen Erkrankungen zu leiden und prinzipiell mit der Teilnahme an einem fMRT Experiment einverstanden zu sein, wurden von uns telefonisch oder per E-Mail kontaktiert. Alle Probanden waren Studierende der Universität Frankfurt am Main, vorwiegend in den Studiengängen Psychologie oder Medizin, und verfügten über eine normale oder zur Normalsichtigkeit korrigierte Sehstärke. Die Studienteilnahme wurde finanziell entschädigt. Die Probanden bekamen die Summe aus dem durchschnittlich von Spieler A zugeteilten Betrag und den von ihnen nicht vergebenen Strafpunkten im Dictator Game ausgezahlt (M = 13.89 Euro, SD = 3.77 Euro). Eine Auflistung des erspieltem Geldbetrags, des Geschlechts, Alters, Studiengangs und Ausschlusskriteriums jedes einzelnen Probanden befindet sich in Tabelle A-1 im Anhang.

2.2 Versuchsablauf und Versuchsmaterialien

Die fMRT Messungen wurden im Zeitraum von Anfang Juni bis Anfang August 2008 am Brain Imaging Center in Frankfurt am Main durchgeführt und fanden vorwiegend abends ab 18 Uhr statt. Zu Beginn jeder Sitzung haben wir die Probanden ein weiteres Mal über die Methode der fMRT sowie deren Ausschlusskriterien und Vorsichtsmaßnahmen aufgeklärt. Die Probanden wurden gebeten ein standardisiertes MRT Sicherheitsprotokoll (siehe Anhang B-1) auszufüllen, mit welchem sie uns erneut bestätigten, keine Risikofaktoren aufzuweisen. Weiterhin wurden sie über den allgemeinen Zweck der Studie und anschließend mittels einer schriftlichen Instruktion im Detail über den Spielablauf des Dictator Games informiert (siehe Anhang B-4). Daraufhin übten die Probanden das Dictator Game mittels eines kurzen Probelaufs am PC außerhalb des Scanners. Es sollte hiermit sichergestellt werden, dass das Prinzip und die Handhabung des Spiels für jeden Versuchsteilnehmer auch praktisch ersichtlich war. Da wir den Einfluss der Fairnessnorm, als ein soziales Motiv, untersuchen wollten, wurden die Probanden an dieser Stelle noch einmal explizit darauf hingewiesen, dass es sich bei den Spielern A um reale Menschen handelte (siehe Knoch et al., 2006; Sanfey et al., 2003).

Wir platzierten die Probanden komfortabel auf der Liege des MRT-Scanners und fixierten ihre Köpfe mittels flexibler Schaumstoffkissen zur Reduktion von eventuell auftretenden Kopfbewegungen. Es erfolgte ein weiterer Probedurchgang, um zu gewährleisten, dass die Probanden mit der ihnen neuen Umgebung und der *Response-Box*, dem Instrument zur Durchführung der im Dictator Game erforderlichen Handlungen, vertraut werden. Daraufhin wurde mit der anatomischen Aufnahme begonnen. Auf diese folgten zwei funktionelle Messungen, in denen das Dictator Game aus zwei verschiedenen Perspektiven gespielt wurde. Den Probanden wurde das Spiel in pseudorandomisierter Reihenfolge in zwei Blöcken dargeboten. Sie begannen je nach Probandennummer zuerst mit der First Person oder mit der Third Party Bedingung.

Den Abschluss jeder Session bildeten eine ausführliche Aufklärung der Probanden bezüglich der genauen Hintergründe der Studie sowie die Auszahlung des erspielten Geldbetrages.

2.2.1 Fragebogenbatterie

Der von der Abteilung Allgemeine Psychologie II für potentielle Sudienteilnehmer zusammengestellte Fragenkatalog enthielt Skalen und Subskalen von etablierten und publizierten Persön-

lichkeitsfragebögen unter anderem zur Erfassung der Konstrukte *Empathiefähigkeit*, *Ungerechtigkeitssensibilität* sowie der *Einstellung gegenüber negativer Reziprozität*.

Aus dem Saarbrücker Persönlichkeitsfragebogen (SPF; Paulus, 2009), der auf dem Interpersonal Reactivity Index (IRI; Davis, 1980) basiert, waren die Subskalen *Empathie*, *Perspektivenübernahme* und *Fantasie* vertreten. Diese lassen sich zu einer Gesamtskala *Empathiefähigkeit* aufaggregieren und umfassen insgesamt 11 Items. Darüber hinaus war die Subskala *Empathie* der deutschen Version des Impulsiveness-Venturesomeness-Empathy Questionnaire (I7) mit 19 Items beigefügt (Eysenck, Daum, Schugens & Diehl, 1990). Zur Erfassung von sozialem Ungerechtigkeitserleben aus verschiedenen Perspektiven waren zwei Subskalen des Fragebogens Sensitivity to Befallen Injustice (SBI; Schmitt, 1996) enthalten. Die erste Skala erfasst die Sensibilität einer Person für Ungerechtigkeit aus der Perspektive eines Betroffenen; die zweite Skala aus der Perspektive eines Beobachters. Beide Skalen setzen sich aus zehn Items zusammen, welche Aufschluss über die dispositionelle Ungerechtigkeitssensibilität einer Person in den jeweiligen Perspektiven geben sollen.

Die Fragebogenbatterie schloss mit der Subskala *Einstellung zu negativer Reziprozität* aus der deutschen Version des Personal Norm of Reciprocity Questionnaire ab (PNRQ; Perugini et al., 2003). Die Subskala umfasst 8 Items, mit welchen die Handlungsdispositionen in Bezug auf negativ gefärbte Beziehungen, die auf Rache oder Vergeltung ausgelegt sind, erfasst werden sollen.

Den Teilnehmern wurden noch weitere Fragebogenskalen vorgelegt, die für andere Studien relevant waren, jedoch in der aktuellen Untersuchung keine Rolle spielten. Der gesamte Fragebogen ist dem Anhang beigefügt (siehe Anhang B-1). Eine Probandin verweigerte das Ausfüllen der Fragebogenskalen aufgrund von moralischen Bedenken, so dass für die differentialpsychologische Auswertung nur Fragebogendaten von 23 Versuchspersonen vorlagen.

2.2.2 Das Dictator Game mit Bestrafungsoption und Fairnessurteil

Die Probanden befanden sich während des Experiments in der Position eines Spielers B und hatten die Aufgabe, die Aufteilung eines Geldbetrages von 20 Euro durch einen realen Spieler A zu beurteilen. Je nach Bedingung war Spieler B selbst von der Zuteilung betroffen (First Person Bedingung) oder befand sich in der Position eines Beobachters, welcher die Zuteilung von ei-

nem Spieler A an einen anderen Spieler C lediglich aus der „dritten Person“ überwachte (Third Party Bedingung). Spieler A wird in diesem Kontext als Dictator bezeichnet, da weder Spieler B noch Spieler C einen Einfluss auf die Betragsaufteilung der 20 Euro hatten. Dem Probanden (Spieler B) wurde jedoch in jedem Durchgang die Möglichkeit gegeben, entweder die subjektive Fairness der Zuteilung zu bewerten (Fairnessdurchgang) oder Strafpunkte zu verteilen (Bestrafungsdurchgang). Pro Bestrafungsdurchgang standen Spieler B vier Strafpunkte im Wert von jeweils 10 Cent zur Verfügung. Jeder nicht vergebene Strafpunkt wurde dem Probanden am Ende des Spieles ausgezahlt. Jeder vergebene Strafpunkt durch Spieler B führte dazu, dass der Geldbetrag, welchen Spieler A für sich behalten wollte, um jeweils 2.50 Euro gekürzt wurde. Somit war der Einsatz von Strafpunkten für beide Partner mit materiellen Kosten verbunden. Die Beurteilung der Fairness einer Zuteilung auf einer Skala von -2 (*sehr unfair*) bis 2 (*sehr fair*) hatte hingegen keine finanziellen Auswirkungen. Das Paradigma ist in Abbildung 4 grafisch veranschaulicht.

Die Bedingungen First Person beziehungsweise Third Party Perspektive des Spielers B wurden nacheinander im Block dargeboten, um den Probanden das Einlassen auf die jeweilige Situation zu erleichtern. Innerhalb jedes Blockes wurden insgesamt 60 Zuteilungen der Spieler A präsentiert. Die Höhe der Zuteilungen sowie die Handlungsoptionen (Fairnessbewertung beziehungsweise Bestrafungsmöglichkeit) war hierbei pseudorandomisiert, um einer Erwartungsbildung oder Ermüdungseffekten der Probanden entgegenzuwirken. Innerhalb eines Blocks sollte

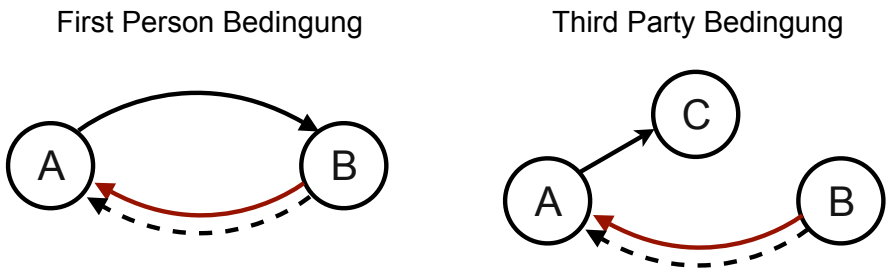

Abbildung 4: First Person und Third Party Bedingung im Dictator Game. In der First Person Bedingung war Spieler B von der Zuteilung (schwarzer Pfeil) direkt betroffen. In der Third Party Bedingung befand sich Spieler B in der Beobachterrolle einer Zuteilung von Spieler A zu einem weiteren Spieler C. In beiden Bedingungen bestand jedoch die Möglichkeit für Spieler B auf die Zuteilung mit einer bestimmten Aktion (Strafpunkte, roter Pfeil oder Fairnessurteil, gestrichelter Pfeil) zu reagieren.

insgesamt 30 mal die Fairness einer Zuteilung beurteilt werden; 30 mal konnten die Probanden Spieler A für eine Zuteilung bestrafen.

Die präsentierten Zuteilungen entstammten einer Auswahl der Angaben von 85 Studenten der Universität Frankfurt. Sie hatten sich während einer Psychologie-Veranstaltung dazu bereiterklärt aus der Sicht eines Dictators zu entscheiden, wie sie jeweils einen Betrag von 20 Euro zwischen sich selbst und zwei realen, ihnen aber völlig unbekannten Personen, aufteilen würden. Sie sollten dabei beachten, dass ihr Gegenüber zum einen ein Spieler B sein würde, welcher durch die Zuteilung direkt betroffen wäre und diese mit einem Abzug von bis zu 10 Euro bestrafen könnte. Zum anderen würde die Zuteilung einen passiven Spieler C betreffen, welchen Spieler B lediglich beobachten würde, doch darauf ebenfalls mit Strafpunkten reagieren könnte. Die 60 in das Dictator Game implementierten Zuteilungen verteilten sich schließlich folgendermaßen: zwanzig 20:0-, zwei 19:1-, zwei 18:2-, zwei 17:3-, vier 15:5-, zwei 13:7-, vier 12:8-, zwei 11:9- sowie zweiundzwanzig 10:10-Zuteilungen. Diese Auswahl orientierte sich an jener der Studie von Strobel et al. (in prep.).

2.3 Technische Umsetzung

2.3.1 Visuelle Stimulation und Response-Box

Die visuelle Stimulation wurde über einen außerhalb des Scanners angebrachten Beamer (Sony, Typ: VPL-XP20) mittels Rückprojektion auf einen sich in der Kopfspule befindlichen Spiegel dargeboten. Jede der 60 Zuteilungen wurde für 2000 ms in Zahlen sowie graphischer Veranschaulichung präsentiert. Die Probanden hatten im Anschluss maximal 6000 ms Zeit mittels einer Response-Box zu reagieren. Durch diese konnten sie mit zwei Tasten auf einer Skala von 0 bis 4 bei Bestrafungsdurchgängen beziehungsweise auf einer Skala von -2 bis 2 bei Fairnessdurchgängen navigieren und mit einer dritten Taste die auf dem Bildschirm aktive Position *einloggen*. Unmittelbar nach der Entscheidung folgte ein Feedback-Bildschirm von mindestens 2000 ms, durch welchen der endgültige finanzielle Stand von Spieler A und Spieler B grafisch anhand von entsprechend modifizierten Balken dargestellt wurde. Das abschließende *Intertrial Intervall*, in dem ein weißes Fixationskreuz auf schwarzem Grund präsentiert wurde, stellte in seiner Länge von durchschnittlich $M = 9000$ ms ($SE = 118$ ms) einen Kompromiss dahingehend dar, einerseits der Trägheit der BOLD-Response gerecht zu werden und andererseits möglichst viele Durchgän-

ge innerhalb eines zeitlich vertretbaren Rahmens zu integrieren (Dale, 1999; Friston, Jezzard & Turner, 1994; Serences, 2004). Insgesamt dauerte ein Durchgang durchschnittlich 5600 ms (siehe Abbildung 5).

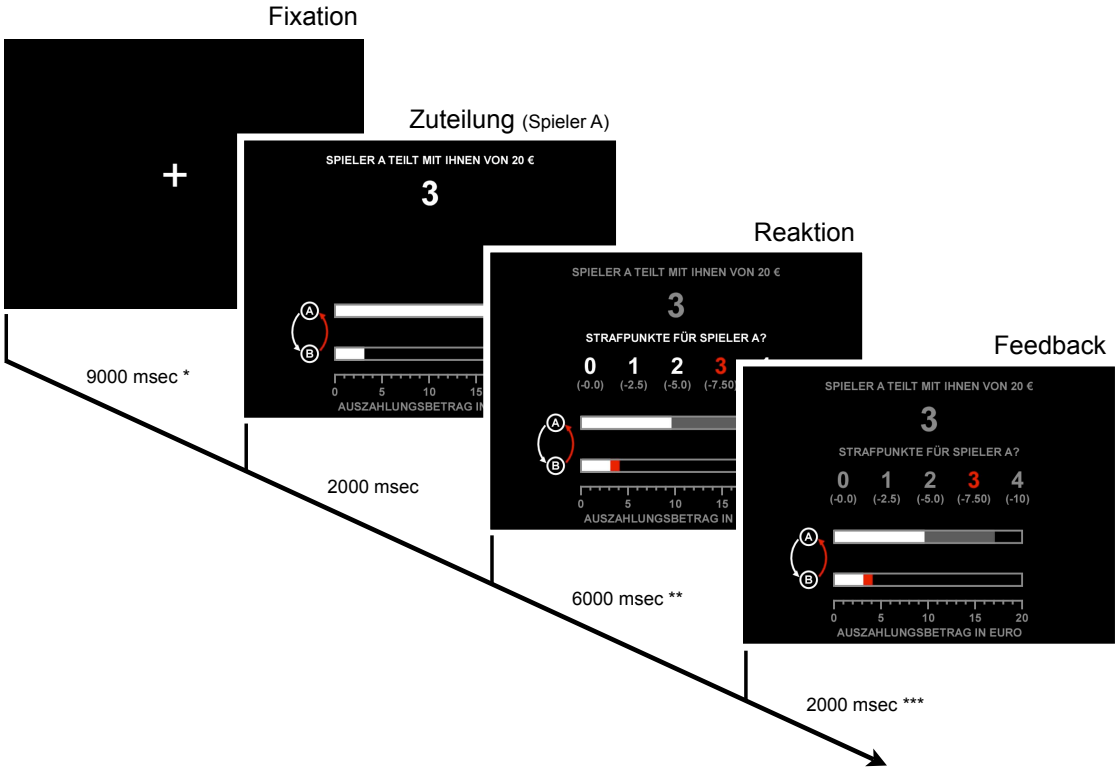

Abbildung 5: Ein 17:3-Durchgang des Dictator Games mit Bestrafungsoption in der First Person Bedingung. Dargestellt ist der zeitliche Ablauf der vier Bildschirmpräsentationen innerhalb des Durchgangs: Fixationskreuz (9000 ms*), Präsentation eines Angebots von Spieler A (2000 ms), Reaktion des Spielers B (6000 ms**) sowie Feedback (2000 ms***). Entscheidet sich Spieler B bei der Zuteilung von 3 Euro durch Spieler A für die Vergabe von 3 Strafpunkten (wie oben abgebildet), hat dies einen Abzug von 30 Cent für Spieler B sowie von 7.50 Euro für Spieler A als Konsequenz.

* 9000 ms *gejittert* (SE = 118 ms)
** 6000 ms als maximale Entscheidungszeit für Spieler B. Tatsächlich loggten die Probanden bereits im Durchschnitt von M = 3424 ms (SE = 82 ms) ein
*** 2000 ms plus 6000 ms minus der jeweiligen Entscheidungszeit (siehe **)

2.3.2 Imaging Parameter

Die Bilder wurden mit einem Siemens Drei-Tesla-Scanner (Allegra, Siemens Medical Systems, Erlangen, Deutschland) und einer Standard Vier-Kanal-Kopfspule aufgenommen. Die T2* gewichteten funktionellen Bilder wurden mit einer *Echo Planar Imaging Sequenz* (EPI) in 36 axialen schiefen Schichten (Schichtauflösung: 3 x 3 mm^2, Schichtdicke: 3 mm, Schichtabstand: 0.75 mm) in absteigender Reihenfolge erhoben (Repetionszeit (TR): 2000 ms, Echozeit (TE): 30 ms, Matrixgröße: 64 x 64). Beide funktionalen Blöcke (First Person und Third Party) bestanden jeweils aus 577 Volumen. Die ersten vier Volumen wurden von der weiteren Analyse entfernt, um T1 Sättigungseffekten entgegenzuwirken. Die hochaufgelösten T1 gewichteten anatomischen dreidimensionalen Bilder wurden mit einer *Magnetization Prepared Rapid Acquisition Gradient Echo Sequenz* (MP-RAGE) in 160 Schichten aufgenommen (TR: 2250 ms, TE: 3.9 ms, Auflösung: 1 x 1 x 1.1 mm^3, Matrixgröße: 256 x 244).

2.4 Statistische Analyse

Dem in der vorliegenden Studie benutzten Dictator Game-Paradigma lag ein zweifaktorieller within-subject Versuchsplan mit Messwiederholung in den Faktoren Perspektive (First Person vs. Third Party) und Handlungsoption (Bestrafung vs. Fairness) zu Grunde.

2.4.1 Verhaltensebene

Die Verhaltensdaten wurden mit dem Statistikprogramm R (R Development Team, siehe www.r-project.org) und dem Software-Paket SPSS (SPSS Inc., Chicago, USA) aufbereitet und analysiert. Unter Verwendung von regressions- und varianzanalytischen Methoden wurden Analysen sowohl auf Individual- und Gruppenebene in Bezug auf die Zuteilungen der Dictator als auch hinsichtlich der Spielerperspektive der Probanden vorgenommen.

Als Reaktionszeit wurde die durchschnittliche Dauer aller Probanden von Beginn der Angebotspräsentation bis zur ersten Reaktion definiert. Die Betrachtung des ersten Tastendrucks zur Begrenzung der Reaktionszeit erschien valider zur Abschätzung der Dauer des Entscheidungsprozesses als die Betrachtung des letzten Tastendrucks zum *Login*. Vor allem bei der Vergabe einer hohen Anzahl von Strafpunkten wäre eine Reaktionszeit, ermittelt über die Login-Taste, stark mit der rein mechanischen Anforderung hierfür konfundiert. Je nach Hypothese wurde

die durchschnittliche Reaktionszeit aller Probanden entweder über die Perspektive oder über die Handlungsoption hinweg gemittelt.

2.4.2 Neuronale Ebene

2.4.2.1 Preprocessing

Die Analyse der fMRT-Daten erfolgte mit dem Software-Paket Brain Voyager QX (Brain Innovation, Maastricht, Niederlande). Die anatomischen Messungen wurden auf eine Voxelgröße von 1 mm^3 *isotropisch* skaliert, um die spätere Koregistrierung von anatomischen und ebenfalls isotropisch formatierten funktionellen Bildern zu vereinfachen. Bei der Aufbereitung der Daten, dem *Preprocessing*, wurden zunächst kleine, während den Messungen aufgetretene Kopfbewegungen korrigiert. Hierbei wurde jedes aufgenommene Volumen eines funktionalen Messblocks durch Translation in den drei Raumrichtungen und Rotation um die drei Raumachsen auf das erste Volumen des Blocks reorientiert (*Six Parameter Rigid Body Motion Correction*). In der anschließenden *Slice Time Correction* wurde eine kubische Spline-Interpolation durchgeführt. Dadurch konnten mögliche Signalverschiebungen durch die zeitlich versetzt aufgenommenen Schichten eines EPI-Volumens korrigiert werden. Um zufällige Ausreißer zu kompensieren und somit das allgemeine Rauschniveau zu verkleinern, wurden die funktionalen Daten weiterhin einem räumlichen *Smoothing* mit einem *Gaußschem Kernel* von 6 mm FwHm (full-width half-maximum) unterzogen. Durch die Verrechnung des Bildgrauwertes jedes Voxels mit dem der benachbarten Voxel konnten zudem interindividuelle funktionale sowie anatomische Variationen bereits in Bezug auf die spätere Gruppenanalyse geglättet werden. Es folgte ein Hochpassfilter zur Beseitigung niedrigfrequenter Drifts und ein Tiefpassfilter zur Reduzierung von hochfrequentem Rauschen im fMRT-Signal, da jene Grenzbereiche des Signals viele Artefakte beinhalten, welche nicht neuronalen Ursprungs sind (*Temporal Filtering*). Nun konnten die einzelnen funktionellen Bilder mit der jeweiligen hochaufgelösten anatomischen Messung jedes Probanden koregistriert, die Anatomien in die AC-PC-Ebene rotiert sowie in den *Talairach-Raum* (Talairach & Tournoux, 1988) transformiert werden. Diese Transformation war notwendig, um die Aktivitätsmuster mehrerer Probanden mit unterschiedlichen Kopfgrößen und -formen in einer Gruppenanalyse vergleichen zu können. Aus den Parametern von Koregistrierung und Normalisierung wurde schließlich ein *Volume Time Course* (VTC), also ein vierdimensionales Datenformat (x, y, z, Zeit), für jeden einzelnen Probanden erstellt.

2.4.2.2 First-Level Analyse

Bei der statistischen Analyse der neuronalen Ergebnisse wurde der Faktor Handlungsoption in vier Handlungsarten aufgegliedert: Bestrafungsverhalten (Vergabe von Strafpunkten), kein Bestrafungsverhalten (keine Vergabe von Strafpunkten), positive Fairnessbeurteilung (Fairnessbewertung von 1 oder 2) und negative Fairnessbeurteilung (Fairnessbewertung von -1 oder -2). Wir erhielten somit insgesamt acht Prädiktoren: First Person-Bestrafung, First Person-keine-Bestrafung, First Person-Fair, First Person-Unfair, Third Party-Bestrafung, Third Party-keine-Bestrafung, Third Party-Fair sowie Third Party-Unfair. Da nur die relativen Schwankungen zwischen den Bedingungen, nicht aber die jeweiligen Mittelwerte selbst interessierten, wurde der Mittelwert der gesamten Zeitreihe im Modell durch eine Konstante angepasst. Als Onset-Parameter der Prädiktoren wurden die individuellen Reaktionszeiten jedes Probanden mit einer selbstgeschriebenen Software in Matlab (The Mathworks, Natick, USA) ermittelt. Die Schätzung der Beta-Gewichte im allgemeinen linearen Modell (ALM) erfolgte nach der Methode der kleinsten Quadrate. Die resultierende Designmatrix ist in Abbildung 6 exemplarisch für einen Probanden grafisch dargestellt.

2.4.2.3 Second-Level Analyse

Auf der Zweiten Ebene (Gruppenanalyse) wurden *Whole Brain Kontraste* berechnet, die es erlauben über das gesamte Gehirn die relativen neuronalen Aktivierungen zwischen den Bedingungen mittels t- beziehungsweise F-Statistiken zu vergleichen. Dabei wurde ein lineares Modell an die Daten angepasst, bei dem die Beta-Gewichte im Gegensatz zur First-Level Analyse frei variieren konnten, um eine adäquate Schätzung auf Populationsebene zu erhalten (zum Random-Effects Modell siehe Holmes & Friston, 1998).

Aufgrund der Problematik des multiplen Testens bei Whole Brain Kontrasten wurden alle Aktivitätsmuster mit dem *FDR-Verfahren* auf dem $p = .01$ Niveau korrigiert, um die Wahrscheinlichkeit des Auftretens von falsch positiven Voxeln zu minimieren (zum False Discovery Rate-Verfahren siehe Genovese, Lazar & Nichols, 2002).

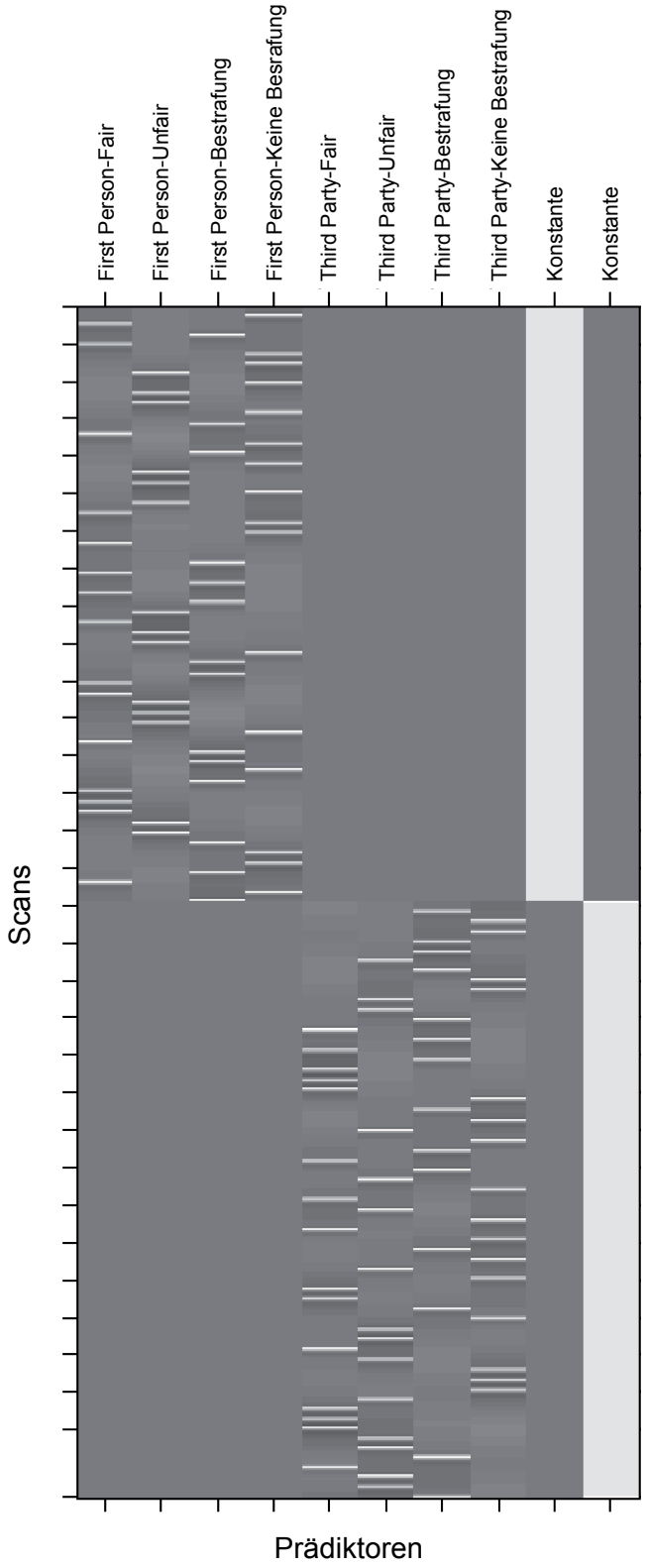

Abbildung 6: Exemplarische Darstellung der Designmatrix im Dictator Game Paradigma des ersten Probanden. Nach der ersten Hälfte der fMRT Scans erfolgte ein Perspektivenwechsel von First Person zu Third Party. Die ersten acht Spalten repräsentieren die acht Prädiktoren, die letzten zwei Spalten veranschaulichen die Konstante im jeweiligen Messblock. Die Stärke des Grauwerts gibt über die Höhe der hämodynamischen Antwortreaktion Auskunft. Die hellsten Bereiche symbolisieren die Peaks des Signals.

2.4.2.4 Event-Related Averages

Neben der Second-Level Analyse auf Grundlage des ALMs wurden in einem weiteren Analyseschritt ereignisbezogene Mittelwerte (*Event-Related Averaging*) berechnet. Dabei wurde die neuronale Aktivität beziehungsweise die hämodynamische Antwortreaktion (siehe 1.4.3) in einem Hirnareal über eine Zeitspanne, in der ein kritisches Verhalten gezeigt wurde (zum Beispiel die Vergabe von Strafpunkten), relativ zur Ruhebedingung, also der Präsentation des Fixationskreuzes, bestimmt. Somit konnten die durchschnittlichen neuronalen Aktivitätsunterschiede aller Probanden in Bezug auf die Handlungsoption und Handlungsperspektive auf Ebene einzelner Hirnregionen detaillierter analysiert werden. Die Auswahl der Hirnregionen erfolgte nach anatomischen Merkmalen auf Grundlage der Whole Brain Kontrastbilder (siehe 2.4.2.3).

Der Zeitraum, über den die durchschnittliche neuronale Aktivität aller Probanden errechnet wurde, orientierte sich an dem erwarteten BOLD-Verlauf (siehe z.B. Boynton, Engel, Glover & Heeger, 1996). Es wurde von der siebten bis zur dreizehnten Sekunde nach der ersten Reaktion gemittelt. Somit konnte der Bereich um den Peak der BOLD-Antwort, von welchem auf das Maxima der Aktivität einer Hirnregion geschlossen werden kann, erfasst werden. Dieser Zeitraum wurde für jedes separate Hirnareal konstant gehalten. Die detaillierte Darstellung des BOLD-Verlaufs je nach Areal und Bedingung kann dem Anhang entnommen werden (siehe Anhang C-2).

2.4.3 Differentialpsychologische Analyse

Neben den allgemeinpsychologisch orientierten Analysen wurden weiterhin differentialpsychologische Aspekte zur möglichen Erklärung von interindividuellen Unterschieden auf behavioraler als auch neuronaler Ebene betrachtet (siehe 2.2.1). Es wurde die Relevanz der Persönlichkeitsmerkmale Ungerechtigkeitssensibilität, Empathie und negative Reziprozität auf die Ausprägung der altruistischen Bestrafung untersucht.

Dabei wurden die Fragebogendaten nach einer Itemanalyse, bei der trennunscharfe Items ausgesondert wurden, skalenweise aggregiert und mit Verhaltenskennwerten sowie mit der neuronalen Aktivität (relativ zur Ruhebedingung betrachtet) bei einem bestimmten Verhalten (beispielsweise der Vergabe von Strafpunkten) korreliert. Die neuronalen Aktivitätswerte wurden in Anlehnung an die Analyse der *Event-Related Averages* berechnet, indem die relative Aktivität

der Voxel von anatomisch definierten Arealclustern gemittelt wurden. Dieses Verfahren führte zu individuellen Aktivitätswerten, die nicht an die Aktivität einzelner Voxel gekoppelt waren. Somit konnten Scheinkorrelationen durch aktivitätsbezogene Selektion und der damit einhergehenden Verletzung der Unabhängigkeit von Messungen vermieden werden (siehe Vul, Harris, Winkielman & Pashler, 2009). Um darüber hinaus Zufallsbefunde zu vermeiden, die zwangsläufig bei der Durchführung von multiplen, voneinander indirekt abhängigen Tests entstehen, wurde streng hypothesen- und konstruktgeleitet vorgegangen und die inferenzstatistischen Ergebnisse mit dem FDR Verfahren korrigiert. Eine vollständige Korrelationstabelle ist dem Anhang beigefügt (siehe Anhang C-2).

3 Ergebnisse

3.1 Verhaltensebene

Im Mittel vergaben die Versuchspersonen einen der vier Strafpunkte, die pro Bestrafungsdurchgang zur Verfügung standen. Die Höhe der Bestrafung schwankte allerdings zwischen den Durchgängen stark. Etwa die Hälfte der Probanden nutzte die gesamte Spannweite der Bestrafungsskala während des Experiments. Die durchschnittliche Variation lag bei SD = 1.28 Strafpunkten (zum Überblick des individuellen Bestrafungsverhaltens siehe Anhang C-1). Die Streuung der investierten Strafpunkte zwischen den Durchgängen ließ sich auf Verhaltensebene insbesondere mit der Fairnesseinschätzung der Probanden und der Angebotshöhe des Spielers A in Zusammenhang bringen.

Die Korrelation der Fairnessbewertungen mit dem Bestrafungsausmaß war bei allen Versuchspersonen erwartungsgemäß negativ und variierte von r = -.42 bis r = -.96. Der Median lag bei r = -.77 (Q_1 = -.70, Q_3 = -.87). Durchschnittlich 58.7% der Variation der Strafpunkte konnte demnach durch die Variation in den Fairnessurteilen erklärt werden.

Neben der wahrgenommenen Fairness war die Angebotshöhe ein entscheidender Faktor bei der Erklärung der Anzahl an vergebenen Strafpunkten. In Abbildung 7 ist die Veränderung der investierten Strafpunkte in Abhängigkeit der Angebotshöhe dargestellt (aggregiert über die Probanden hinweg). Die Vergabe von Strafpunkten stieg annähernd linear mit der Größe des Anteils, den Spielers A für sich beanspruchte. Bei der Aufteilung von 10 Euro für den Probanden und 10 Euro für Spieler A (Equal Split) wurden im Durchschnitt null Strafpunkte (über First Per-

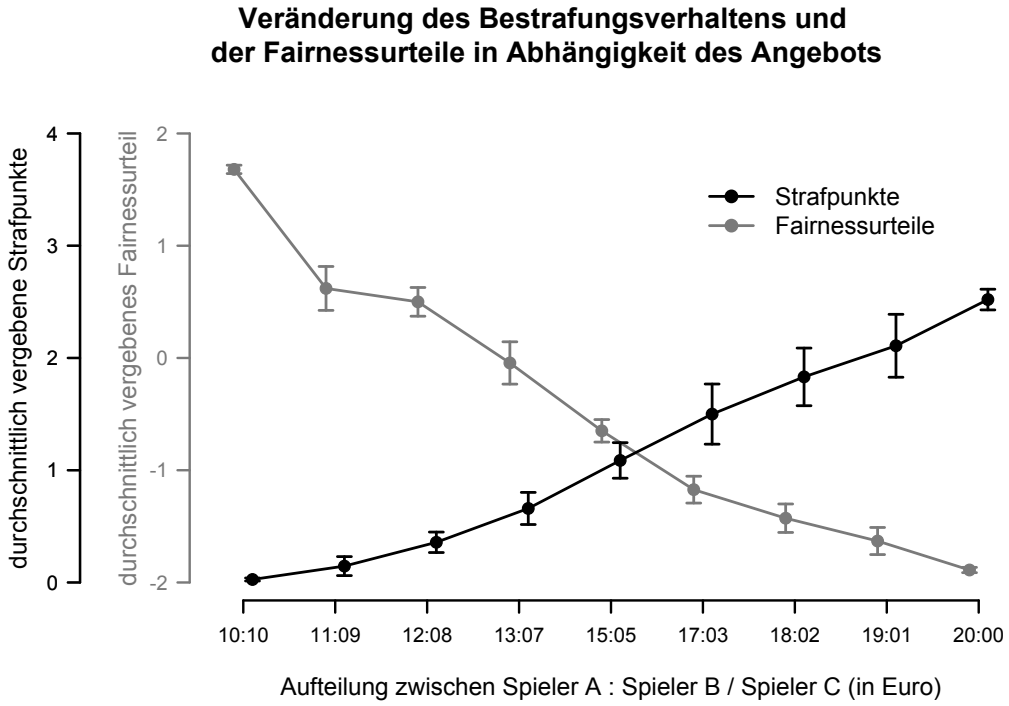

Abbildung 7: Durchschnittlich vergebene Strafpunkte und Fairnessurteile (von -2 = *sehr unfair* bis 2 = *sehr fair*) je Angebot des Spielers A über First Person und Third Party Perspektive gemittelt (Fehlerbalken entsprechen den jeweiligen Standardfehlern).

son und Third Party Bedingung gemittelt) vergeben. Die durchschnittliche Bestrafung stieg auf einen Strafpunkt bei der Aufteilung von 5 Euro für den Probanden und 15 Euro für Spieler A und erreichte eine Höhe von 2.5 Strafpunkten bei der Aufteilung von 0 Euro für den Probanden und 20 Euro für Spieler A. Eine Varianzanalyse mit Messwiederholung auf dem Faktor „Angebot“ (neun Stufen: von 10:10 bis 20:00) zeigte, dass die Veränderung in der Angebotshöhe einen hochsignifikanten Anteil von $\eta^2 = 0.56$ der Varianz der Bestrafungshöhe erklären konnte ($F_{(8,207)} = 220.59$, $p < .001$).

In Abbildung 7 ist darüber hinaus die Veränderung der Fairnessurteile in Abhängigkeit des Angebots abgetragen. Wie erwartet und komplementär zum Bestrafungstrend sank der Durchschnitt der Fairnessurteile mit zunehmend ungleich werdender Aufteilung zu Gunsten von Spieler A. Insgesamt konnte ein Anteil von $\eta^2 = 0.88$ der Fairnessurteilsvarianz durch die Stufen der Angebote erklärt werden ($F_{(8,207)} = 1240.5$, $p < .001$).

3.1.1 Verhaltenskonsistenz

Alle Versuchspersonen des analysierten Datensatzes vergaben Strafpunkte. Die durchschnittliche Bestrafungshöhe variierte jedoch im Experiment zwischen den Probanden hoch. In einem Range von 0.2 bis 2.1 streute der Mittelwert der insgesamt investierten Strafpunkte interindividuell. In einem weiteren Analyseschritt wurde überprüft, ob auch intraindividuell eine Veränderung der durchschnittlichen Strafpunktvergabe stattfand. Dafür wurden Durchschnittswerte je Proband und experimenteller Perspektive (First Person vs. Third Party Bedingung) berechnet. Es zeigte sich, dass die durchschnittliche Bestrafungshöhe zwischen den Bedingungen mit $r = .85$ hoch korrelierte ($t_{(22)} = 7.66$, $p < .001$).

Dennoch gaben die Probanden nicht ausnahmslos gleich viele Strafpunkte in den beiden Bedingungen aus. In der durchgeführten Regressionsanalyse konnte die Verhaltensabweichung durch den Steigungskoeffizienten b ermittelt werden. Bei keiner Verhaltensänderung sollte die Schätzung dieses Parameters den Wert $\hat{b} = 1$ annehmen. Stattdessen war der Steigungskoeffizient der errechneten Regressionsgleichung mit $b = 0.76$ niedriger. In Abbildung 8 ist dies grafisch veranschaulicht. Die theoretische Regressionsgerade unter Annahme keiner Änderung im Verhalten (gestrichelte Linie) liegt außerhalb des Konfidenzintervalls der empirischen Regressionsgeraden (grauer Bereich der durchgezogenen Linie in Abbildung 8). Die empirische Regressionsgerade zeigt im Vergleich zur theoretischen Gerade, dass die Probanden, die relativ wenig Strafpunkte vergaben, dies zu einem größeren Anteil in der First Person Bedingung taten, während Probanden, die generell mehr investierten, dies in der Third Party Bedingung taten. Der Unterschied zwischen erwartetem Steigungskoeffizient und dem empirischen Koeffizienten war statistisch signifikant ($t_{(22)} = -2.45$, $p = .02$), wenngleich der Unterschied als klein einzuschätzen ist und die Konstante der Regressionsgleichung sich nicht bedeutsam von der erwarteten Konstante $\hat{a} = 0$ unterschied ($t_{(22)} = 1.72$, $p = .10$).

Um zu überprüfen, ob mit der Änderung der Spielerperspektive eine Veränderung der Fairnesswahrnehmung einherging, wurden die Fairnessurteile jedes Probanden über die jeweilige experimentelle Bedingung gemittelt. In Abbildung 9 ist der Zusammenhang der Mittelwerte der Fairnessurteile in der First Person Bedingung mit dem Zusammenhang der Mittelwerte in der Third Party Bedingung dargestellt. Die Fairnessurteile korrelierten mit $r = .79$ ähnlich hoch wie die Bestrafungspunkte ($t_{(22)} = 6.10$, $p < .001$). Im Gegensatz zur Analyse des Bestrafungsverhal-

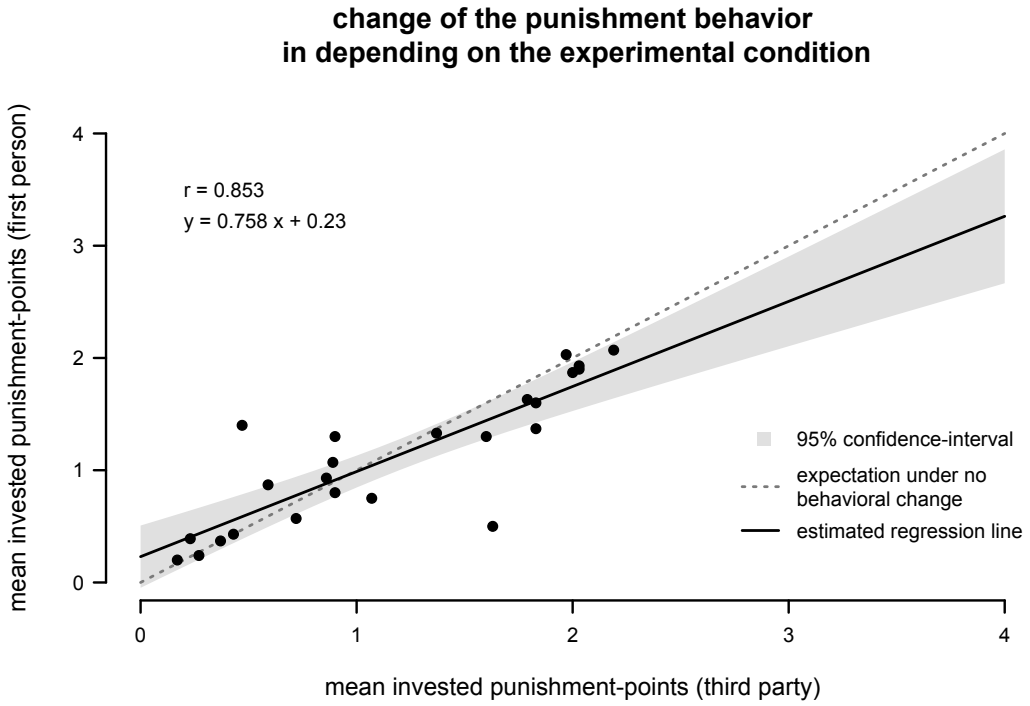

Abbildung 8: Zusammenhang der vergebenen Strafpunkte in der First Person Bedingung mit den vergebenen Strafpunkten in der Third Party Bedingung (Datenpunkte entsprechen dem jeweiligen Durchschnitt der vergebenen Strafpunkte je Proband).

tens zeigten die geschätzten Regressionsparameter keine statistisch bedeutsame Abweichung von den zu erwarteten Parametern unter der Annahme keiner Veränderung in der durchschnittlichen Fairnessbewertung (Konstante a: $t_{(22)} = 0.141$, $p = .89$; Steigung b: $t_{(22)} = -1.735$, $p = .10$). In Abbildung 9 ist dies anhand des Vergleichs der gestrichelten Linie (Regressionsgerade unter Annahme keiner Änderung in den Fairnessurteilen) mit der durchgezogenen Linie (empirische Regressionsgerade) erkennbar. Dieser Befund ist ein Indiz dafür, dass die Probanden die Angebote in der First Person Bedingung im Mittel annähernd ähnlich gerecht oder ungerecht wahrnahmen wie in der Third Party Bedingung. Diese Schlussfolgerung ist allerdings unter der Einschränkung der unbekannten Beta-Irrtumswahrscheinlichkeit zu betrachten.

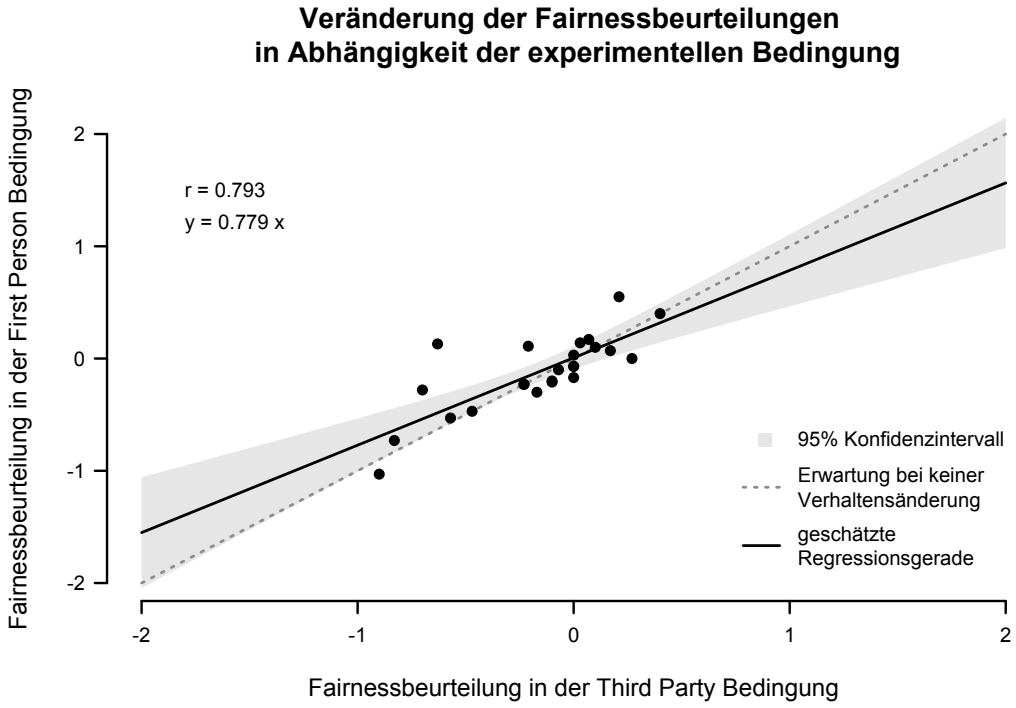

Abbildung 9: Zusammenhang der Fairnessbeurteilungen (von -2 = *sehr unfair* bis 2 = *sehr fair*) in der First Person Bedingung mit den Fairnessbeurteilungen in der Third Party Bedingung (Datenpunkte entsprechen dem jeweiligen Durchschnitt der Fairnessbeurteilungen je Proband).

3.1.2 Reaktionszeiten

In Abbildung 10a ist die durchschnittliche Zeit bis zum ersten Tastendruck für Durchgänge aus First Person Perspektive und Third Party Perspektive als Violinplot dargestellt (zum Violinplot siehe Hintze & Nelson, 1998). Der Median lag in der First Person Bedingung mit $Md = 3.31$ wie erwartet niedriger als in der Third Party Bedingung ($Md = 3.47$). Shapiro-Wilk-Tests zeigten, dass von Normalverteilung der Messwerte nicht ausgegangen werden konnte. Daher wurde zur Überprüfung der statistischen Bedeutsamkeit dieses Unterschieds ein nonparametrisches Verfahren angewendet. Ein Mann-Whitney U-Test ergab, dass ein signifikanter Unterschied in den Reaktionszeiten zwischen den experimentellen Bedingungen im Bezug auf die Reaktionszeiten auf dem $\alpha = .05$-Niveau vorlag ($U = 828.5$, $p = 0.03$, $d = 0.31$ – kleine Effektstärke nach Cohen, 1988). In der Third Party Bedingung brauchten die Probanden demnach signifikant länger bis zum ersten Tastendruck.

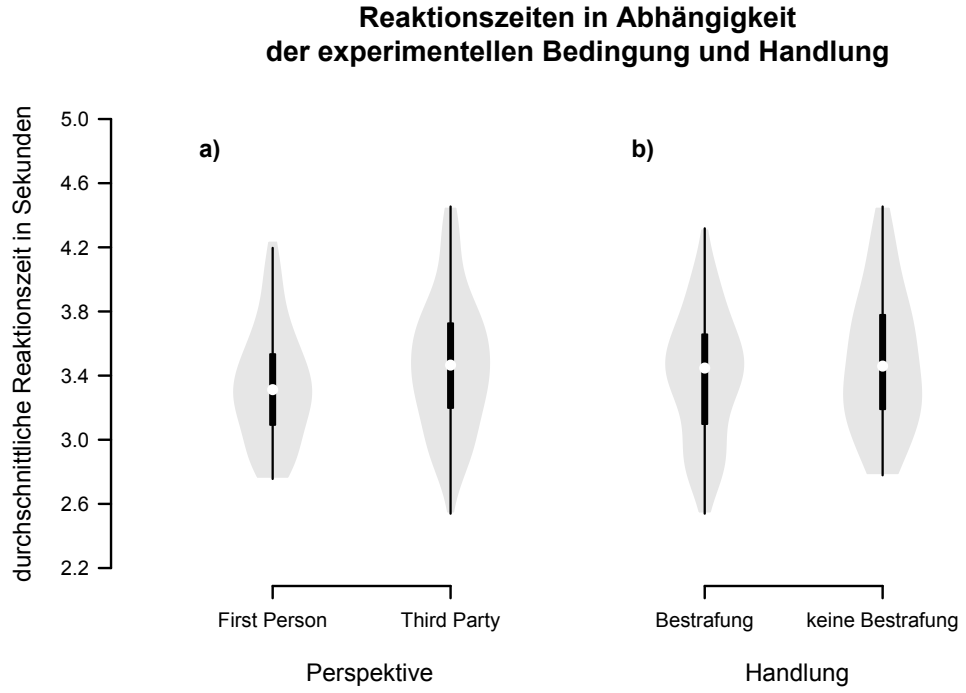

Abbildung 10: Violinplots zum Vergleich der durchschnittlich vergangenen Zeit bis zum ersten Tastendruck, separiert nach der experimentellen Perspektive (a) und der Handlung in Bestrafungsdurchgängen (b).

Dagegen zeigte sich keine signifikante Abweichung in den Reaktionszeiten zwischen den Durchgängen, in denen die Probanden bestraften und den Durchgängen, in denen sie dies nicht taten ($U = 91$, $p = 0.51$ – auch bei diesem Vergleich wurde ein nonparametrisches Verfahren gewählt, da von Normalverteilung der Daten nicht ausgegangen werden konnte). Wie in Abbildung 10b zu erkennen, zeigten die Verteilungen der Reaktionszeiten zwischen den Handlungsarten eine hohe Überschneidung und die Mediane lagen sehr nah beieinander ($Md = 3.45$ für Bestrafungsverhalten gegenüber $Md = 3.46$, wenn keine Strafpunkte vergeben wurden).

3.2 Neuronale Ebene

Auf neuronaler Ebene wurden in Anlehnung an die Studie von Strobel et al. (in prep.) zunächst der Einzelkontrast „Bestrafungsverhalten größer keine Bestrafung“ analysiert. Danach wurden die orthogonalen Hauptkontraste „First Person größer Third Party“ und „Bestrafungsdurchgänge größer Fairnessdurchgänge“ näher betrachtet.

In einem zweiten Analyseschritt wurden mit Hilfe der Event-Related Averages die Aktivitätsmuster auf Arealebene zwischen den Handlungsarten verglichen. Die angegebenen Gehirn-Koordinaten entsprechen dem Koordinatensystem nach Talairach und werden nachfolgend mit Tal abgekürzt (Talairach & Tournoux, 1988).

3.2.1 Second-Level Analyse

3.2.1.1 Einzelkontrast: Bestrafung versus keine Bestrafung

Beim Einzelkontrast „Bestrafung versus keine Bestrafung“ zeigten sich erwartungsgemäß eine höhere neuronale Signalstärke im linken wie rechten DLPFC sowie dem ACC. Auch die Insula war sowohl links- als auch rechtshemisphärisch bei Bestrafungsverhalten signifikant aktiver. Darüber hinaus zeigten Teile des Thalamus eine erhöhte Aktivität, wenn die Probanden sich entschieden Strafpunkte zu vergeben. Entgegen den Befunden von Strobel et al. (in prep.) zeigten sich subkortikal keine Aktivitätsunterschiede in Arealen, die mit Belohnungsantizipation assoziiert werden (wie dem N. caudatus oder dem N. accumbens). Des Weiteren waren motorische Areale des Neocortex bei Bestrafungsverhalten bedeutsam aktiver, was sich durch die größere motorische Aktivität erklären lässt, die mit der Vergabe von Strafpunkten im gewählten Paradigma einherging.

Zur Visualisierung der Aktivitätsunterschiede wurden die statistischen Kennwerte voxelweise auf eine gemittelte anatomische Aufnahme aller Probanden projiziert (siehe Abbildung 11). Die farblich hervorgehobenen Stellen markieren Bereiche im Gehirn, bei denen ein statistisch bedeutsam höheres neuronales Signal gemessen wurde, wenn die Probanden Punkte zur Bestrafung von Person A investierten im Kontrast zu Durchgängen, in denen sie keine Strafpunkte einsetzten.

3.2.1.2 Kontrast: First Person versus Third Party

Auch beim Kontrast „First Person versus Third Party“ konnten entgegen der Erwartung keine Aktivitätsunterschiede im N. accumbens festgestellt werden. Stattdessen waren Teile des dorsalen Striatums bedeutsam aktiver, wenn die Probanden das Dictator Game aus der First Person Perspektive spielten (darunter der N. caudatus sowie Teile des Putamen).

Auf neokortikaler Ebene war wie erwartet der PC in der First Person Bedingung deutlich

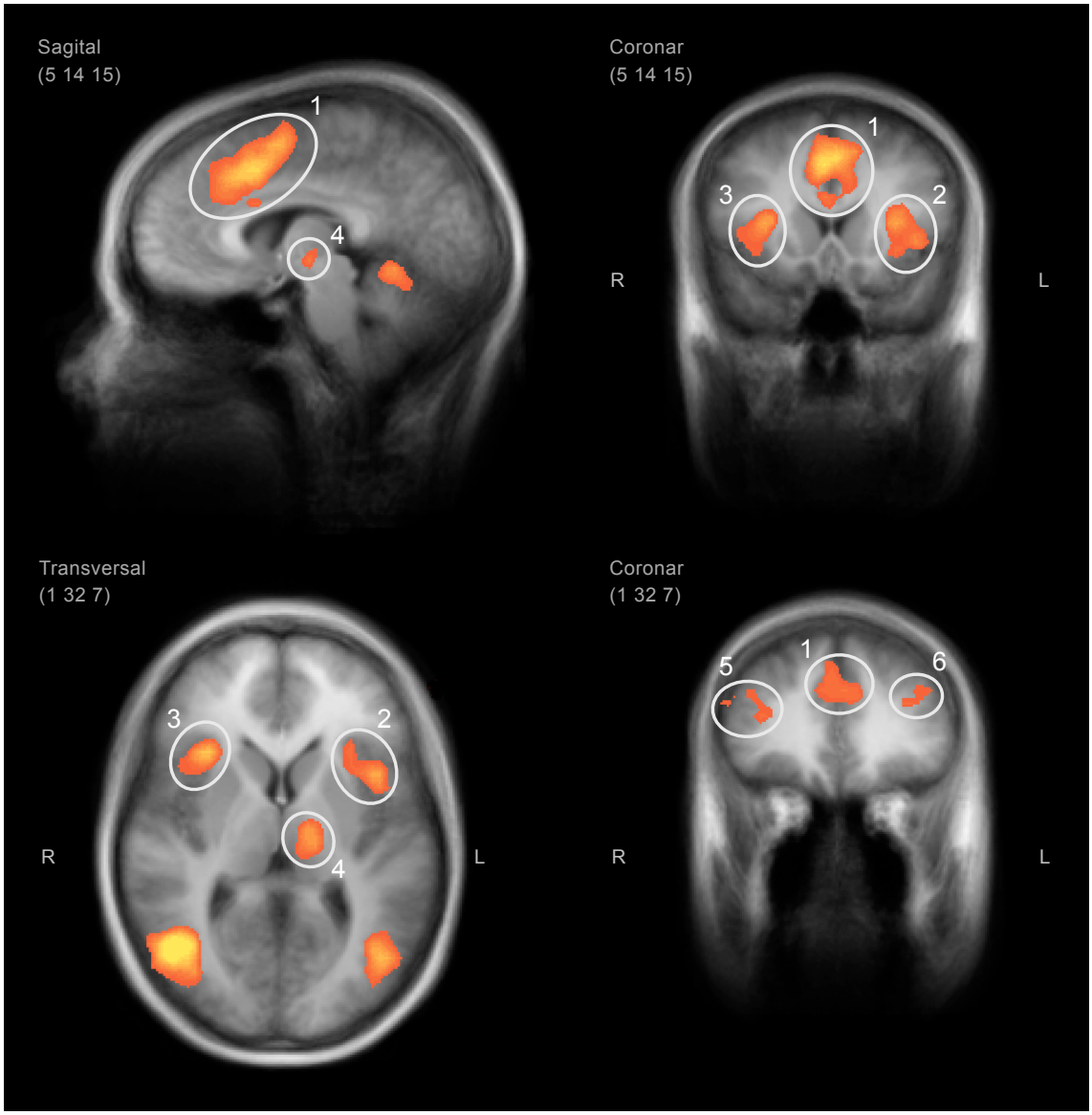

Abbildung 11: Statistische Karte des Kontrasts „Bestrafung größer keine Bestrafung". 1 = anteriorer cingulärer Cortex (ACC), 2 = linke Insula, 3 = rechte Insula, 4 = Thalamus (medial-dorsaler Nucleus), 5 = rechter dorsolateraler präfrontaler Cortex (rDLPFC), 6 = linker dorsolateraler präfrontaler Cortex (lDLPFC). In Klammern: x, y, und z-Koordinaten nach Talairach. $P_{FDR} = .01$.

aktiver. Der Haupteffekt „First Person versus Third Party“ zeigte des Weiteren signifikante Resultate in dem ventralen und dorsalen ACC, dem inferior-parietalen Lobus sowie dem im parietalen Lobus gelegenen Precuneus. Auch die neuronale Signalstärke von Teilen des Thalamus war statistisch bedeutsam höher in der First Person Bedingung. Analog zum Kontrast „Bestrafung versus keine Bestrafung“ war die linke Insula und der linke DLPFC aktiver, wenn die Probanden von den Angeboten des Spielers A direkt betroffen waren im Vergleich zu der Situation, in der die Probanden die Interaktion zwischen Spieler A und einem dritten Spieler C aus der Beobachterrolle heraus verfolgten.

Die aufgeführten Aktivitätsunterschiede des Kontrasts „First Person versus Third Party“ sind in Abbildung 12 grafisch veranschaulicht.

3.2.1.3 Kontrast: Bestrafungsdurchgänge versus Fairnessdurchgänge

Zuletzt wurde das Aktivitätsmuster bei Bestrafungsdurchgängen dem Aktivitätsmuster bei Fairnessdurchgängen gegenübergestellt. Abbildung 13 gibt einen Überblick über die Bereiche des Gehirns, die bei Bestrafungsdurchgängen im Kontrast zu Fairnessdurchgängen eine statistisch bedeutsame Rolle spielten. Der ACC, der linke und rechte DLPFC sowie die rechte Insula waren signifikant aktiver, wenn die Probanden die Möglichkeiten hatten Strafpunkte zu vergeben, als wenn sie lediglich die Fairness eines Angebots beurteilen sollten. Ähnlich wie beim ersten Hauptkontrast „First Person versus Third Party“ zeigte sich auch bei dem Vergleich von Bestrafungsdurchgängen und Fairnessdurchgängen eine signifikante Abweichung der neuronalen Aktivität im Precuneus und dem PC zugunsten der Bestrafungsdurchgänge. In Arealen, die mit Belohnungsantizipation assoziiert sind, wie dem N. caudatus oder dem N. accumbens, zeigten sich dagegen keine bedeutsamen Abweichungen in der Signalintensität. Stattdessen zeigte sich eine signifikant höhere Aktivität im mittleren frontalen Gyrus, einer Hirnregion, über deren Rolle beim Phänomen der altruistischen Bestrafung keine konkreten Hypothesen vorlagen.

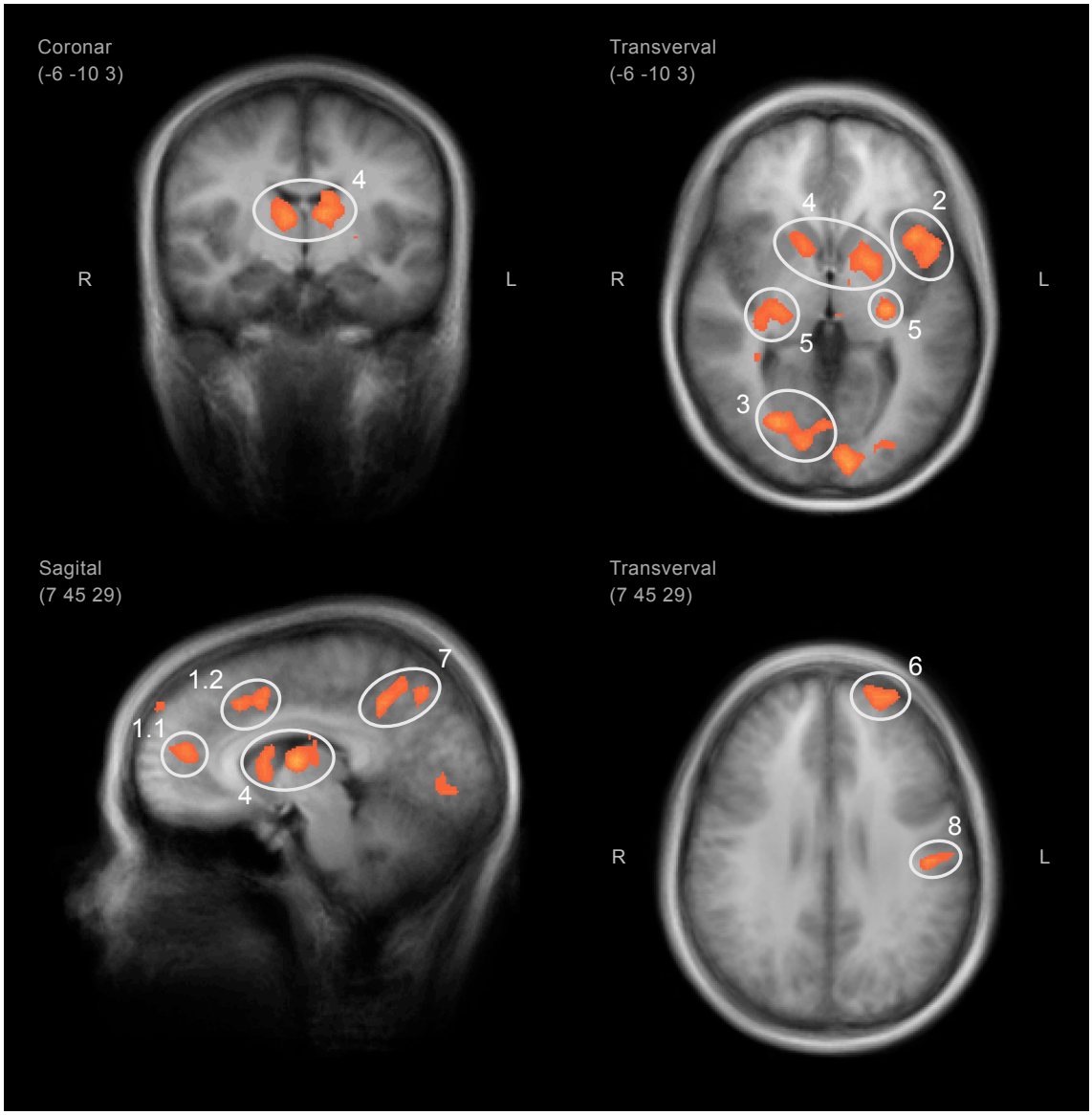

Abbildung 12: Statistische Karte des Kontrasts „First Person größer Third Party". 1 = anteriorer cingulärer Cortex (1.1 = ventraler ACC, 1.2 dorsaler ACC), 2 = linke Insula, 3 = Precuneus, 4 = dorsales Striatum (Nucleus caudatus sowie Teile des Putamen), 5 = Teile des Thalamus, 6 = linker dorsolateraler präfrontaler Cortex (lDLPFC), 7 = posteriorer cingulärer Cortex (PC) sowie Teile des Precuneus, 8 = inferior-parietaler Lobus. In Klammern: x, y, und z-Koordinaten nach Talairach. $P_{FDR} = .01$.

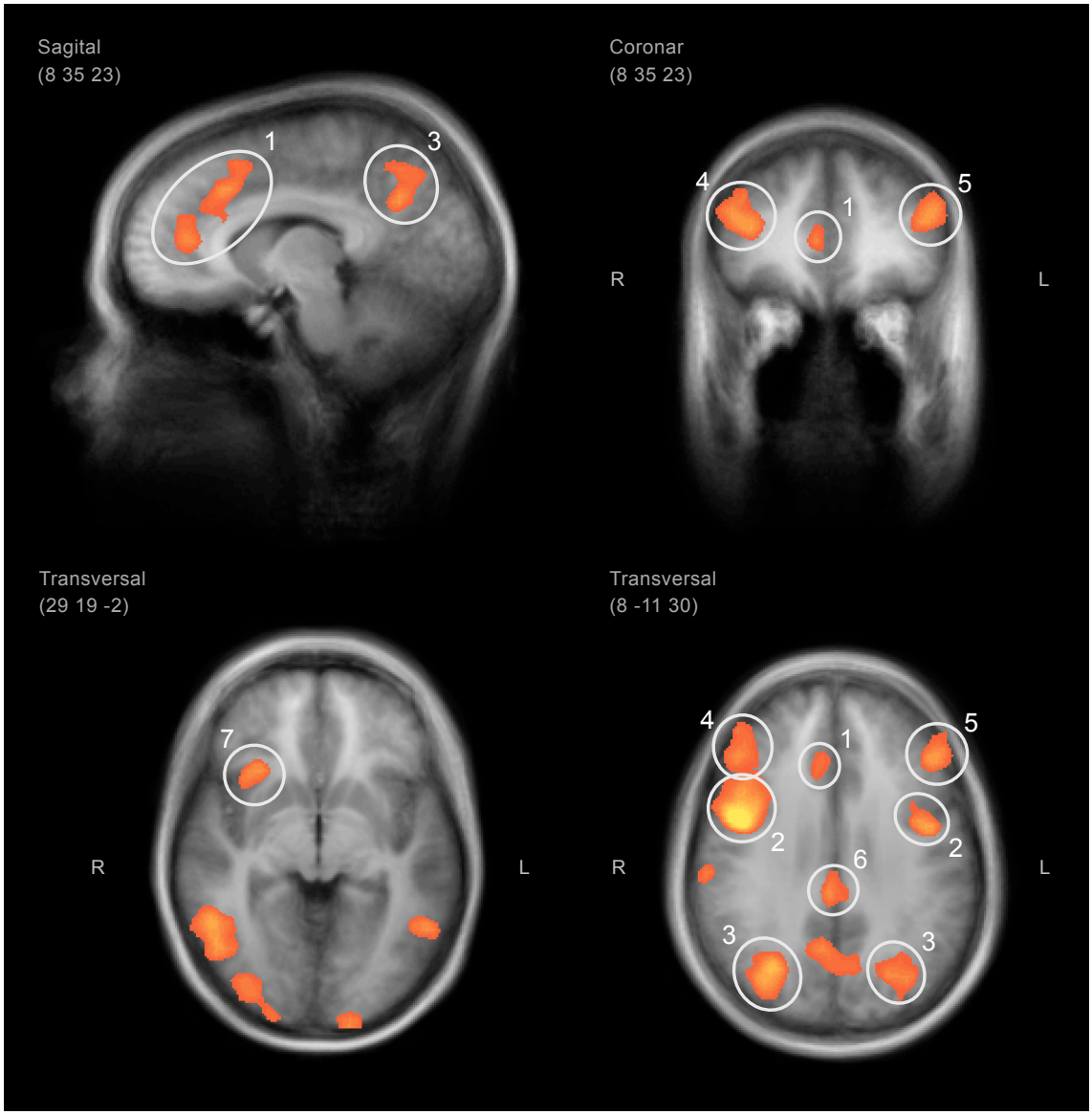

Abbildung 13: Statistische Karte des Kontrasts „Bestrafungsdurchgänge größer Fairnessdurchgänge". 1 = anteriorer cingulärer Cortex (ACC), 2 = mittlerer frontaler Gyrus, 3 = Precuneus, 4 = rechter dorsolateraler präfrontaler Cortex (rDLPFC), 5 = linker dorsolateraler präfrontaler Cortex (lDLPFC), 6 = posteriorer cingulärer Cortex (PC), 7 = rechte Insula. In Klammern: x, y, und z-Koordinaten nach Talairach. $P_{FDR} = .01$.

3.2.2 Ereignisbezogene Aktivitätsunterschiede

Die Analyse der Hauptkontraste ließ keine Differenzierung der Aktivierungsunterschiede zwischen den Einzelbedingungen zu. Allerdings spielten bestimmte Hirnregionen, wie die Insula, der DLPFC oder der Precuneus, bei mehreren Kontrasten eine bedeutsame Rolle. Die Event-Related Averages sollten eventuell vorliegende Interaktionen aufdecken, die durch Aggregation über Bedingungs- oder Handlungsstufen hinweg bei den Hauptkontrasten nicht detektiert werden konnten. Auf Haupteffekte wird im Folgenden nicht mehr eingegangen, da diese bereits durch die Second-Level Analyse hinreichend untersucht wurden. Durch die Event-Related Averages konnte außerdem geklärt werden, ob die signifikanten Aktivitätsdifferenzen durch eine Deaktivierung oder Aktivierung in dem betreffenden Areal in einer Bedingung hervorgerufen wurden (zur negativen BOLD siehe Shmuel, Yacoub, Pfeuffer, de Moortele, Adriany, Hu & Ugurbil, 2002; Wade, 2002).

3.2.2.1 Insula

In Abbildung 14 und Abbildung 15 sind die gemittelten prozentualen Aktivitätsunterschiede der rechten und linken Insula je nach Bedingung, Perspektive und Handlungsart dargestellt.

Deskriptiv war das Aktivitätsmuster in der rechten und linken Insula ähnlich. Während sich bei Fairnessdurchgängen die geringsten Signalveränderungen im Vergleich zur Ruhebedingung zeigten und keine Unterschiede zwischen Perspektive (First Person vs. Third Party) und Handlungsart (positive Fairnessbewertung vs. negative Fairnessbewertung) ausgemacht werden konnten, so waren die höchsten Aktivitätsunterschiede bei Bestrafungshandlungen zu verzeichnen. Insbesondere in der rechten Insula war die Differenz der Signalveränderung zwischen Bestrafung und keiner Bestrafung in der First Person Bedingung weitaus kleiner als die Differenz dieser beiden Handlungsarten in der Third Party Bedingung. Diese Interaktion weist darauf hin, dass die Art der Handlung mit einer unterschiedlichen Aktivitätshöhe in der Insula einherging, wenn die Probanden das Paradigma aus der Beobachterperspektive heraus spielten, während dies nicht (rechte Insula) beziehungsweise nicht so ausgeprägt (linke Insula) der Fall war, wenn die Probanden direkt von den Angeboten betroffen waren.

Eine Varianzanalyse mit den Stufen Bedingung (Bestrafungsdurchgang vs. Fairnessdurchgang) x Perspektive (First Person vs. Third Party) x Bestrafungshandlung (Bestrafung vs. kei-

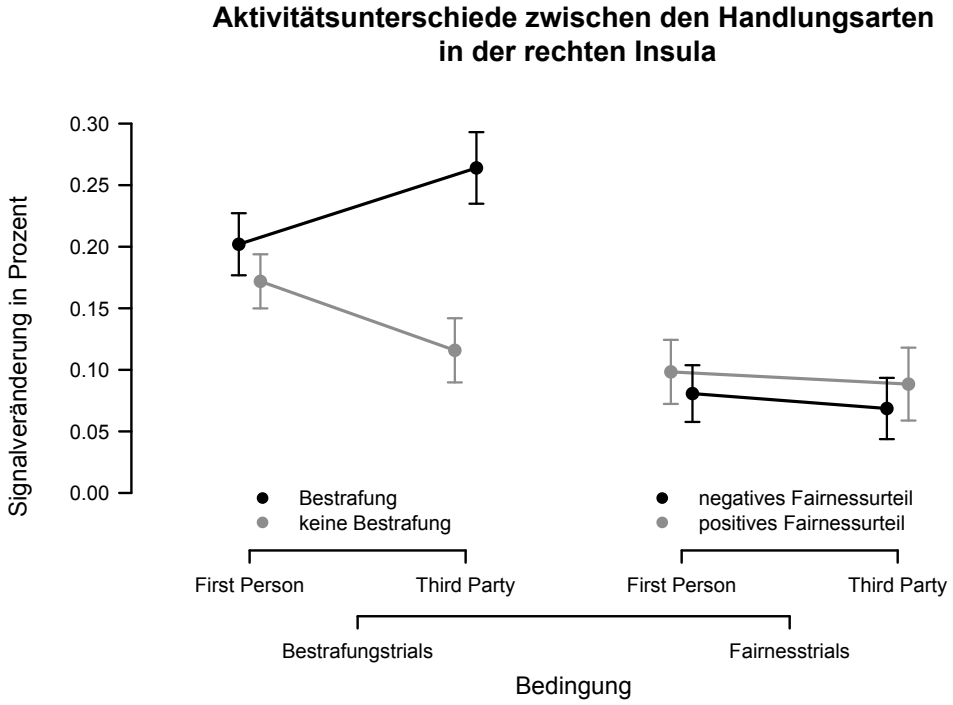

Abbildung 14: Gemittelte prozentuale Aktivitätsunterschiede in der rechten Insula aufgeschlüsselt nach Bedingung, Perspektive und Handlungsarten. Fehlerbalken entsprechen einer Standardfehlerabweichung. Tal.-Koordinaten (x, y, z): 30 23 4.

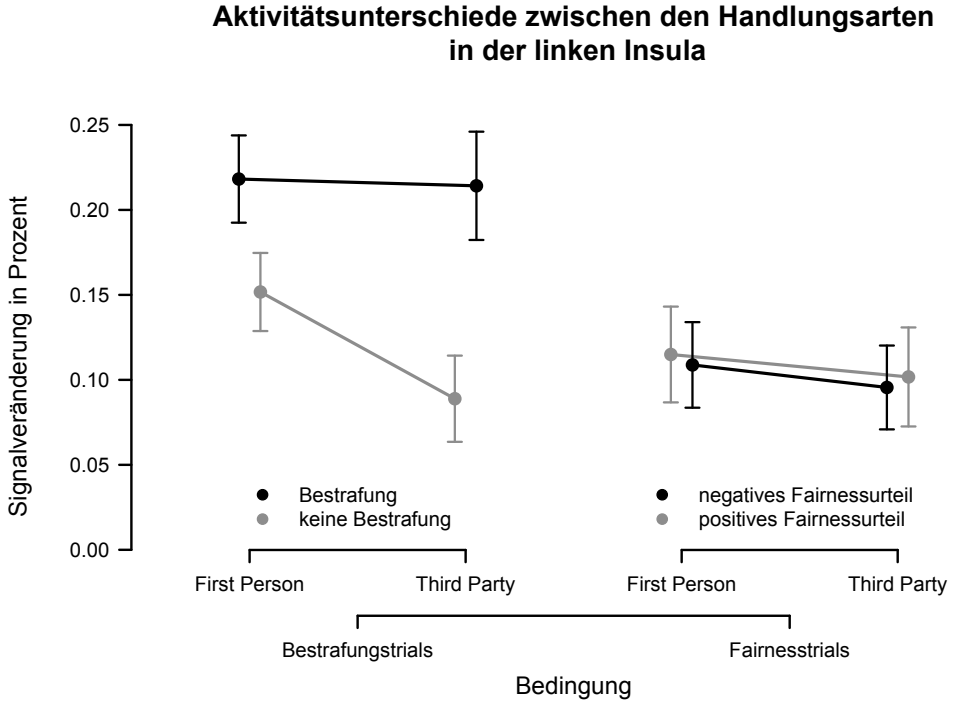

Abbildung 15: Prozentuale Aktivitätsunterschiede in der linken Insula aufgeschlüsselt nach Bedingung, Perspektive und Handlungsarten. Fehlerbalken entsprechen einer Standardfehlerabweichung. Tal.-Koordinaten (x, y, z): -31 16 4.

ne Bestrafung) x Fairnesshandlung (negatives Fairnessurteil vs. positives Fairnessurteil) ergab, dass die Interaktion Perspektive x Bestrafungshandlung für die rechte Insula signifikant war ($F_{(1,46)} = 5.21$, $p = .02$, $\eta^2 = .02$). In der linken Insula war dies nicht der Fall ($F_{(1,46)} = 1.21$, $p = .27$, $\eta^2 < .01$). Dieses Ergebnis erklärt, warum lediglich für die linke Insula beim Hauptkontrast „First Person versus Third Party“ auf Second-Level Analyse ein signifikanter Aktivitätsunterschied nachgewiesen werden konnte.

3.2.2.2 Dorsolateraler präfrontaler Cortex

Auch die Muster der Signalveränderung in dem linken und rechten DLPFC folgten vergleichbaren Trends; die höchsten Aktivitätsunterschiede zeigten sich bei Bestrafungshandlungen. Die Signalveränderungen in den Fairnessdurchgängen waren auch bei dieser Hirnregion generell und unabhängig der Bewertungsart und Perspektive niedrig (siehe Abbildung 16 und 17).

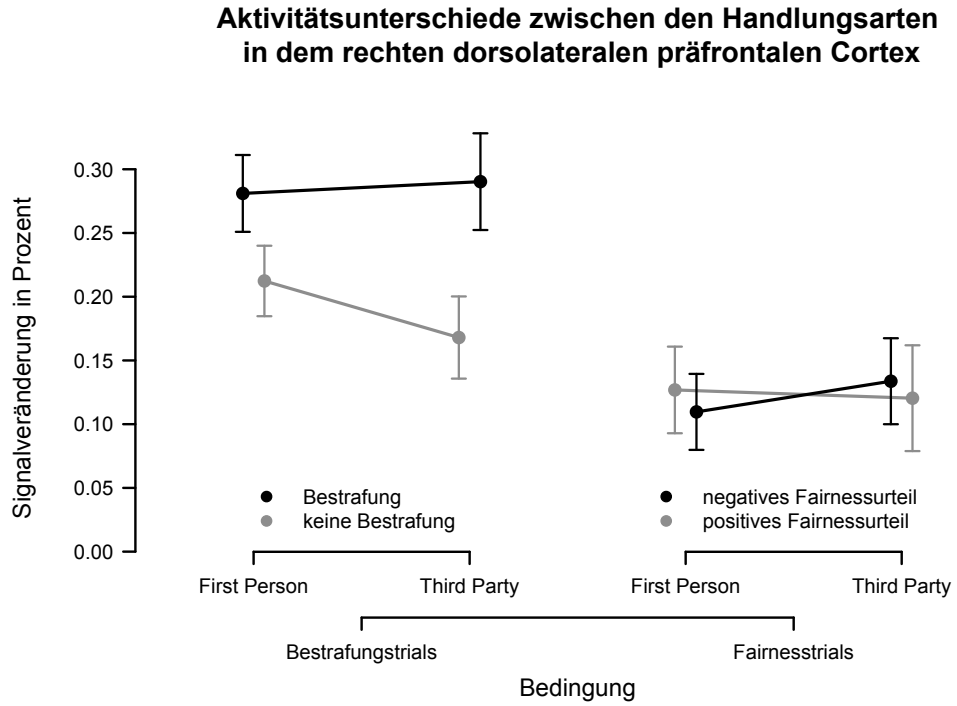

Abbildung 16: Prozentuale Aktivitätsunterschiede in dem rechten dorsolateralem präfrontalem Cortex aufgeschlüsselt nach Bedingung, Perspektive und Handlungsarten. Fehlerbalken entsprechen einer Standardfehlerabweichung. Tal.-Koordinaten (x, y, z): 35 38 25.

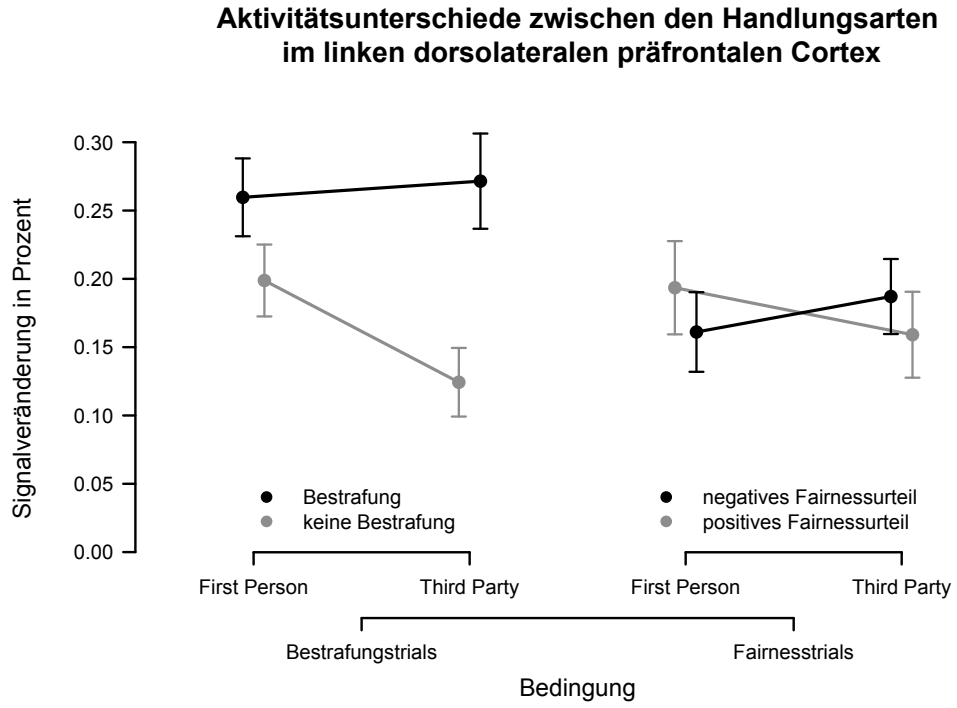

Abbildung 17: Prozentuale Aktivitätsunterschiede in dem linken dorsolateralem präfrontalem Cortex aufgeschlüsselt nach Bedingung, Perspektive und Handlungsarten. Fehlerbalken entsprechen einer Standardfehlerabweichung. Tal.-Koordinaten (x, y, z): -36 30 33.

Wenn Probanden sich in der First Person Perspektive entschieden, keine Strafpunkte zu vergeben, hing dies mit einer Aktivierung des DLPFCs zusammen. Diese Aktivität war höher als bei Probanden, die aus der Third Party Perspektive handelten.

Die deskriptiv erkennbare Interaktion Perspektive x Bestrafungshandlung wurde dagegen weder im rechten noch im linken DLPFC signifikant (rechter DLPFC: $F_{(1,46)} = 0.63$, $p = .43$, $\eta^2 < .01$; linker DLPFC: $F_{(1,46)} = 2.10$, $p = .15$, $\eta^2 = .01$). Lediglich die schon gefundenen Haupteffekte konnten einen statistisch bedeutsamen Anteil der Signalunterschiede erklären.

3.2.2.3 Precuneus

In der Second-Level Analyse zeigte sich ein signifikanter Aktivitätsunterschied des Precuneus im Hauptkontrast „First Person versus Third Party“. Der Trend der Fairnessdurchgänge in den Event-Related Averages war mit diesem Ergebnis konsistent. Unabhängig der Bewertung war

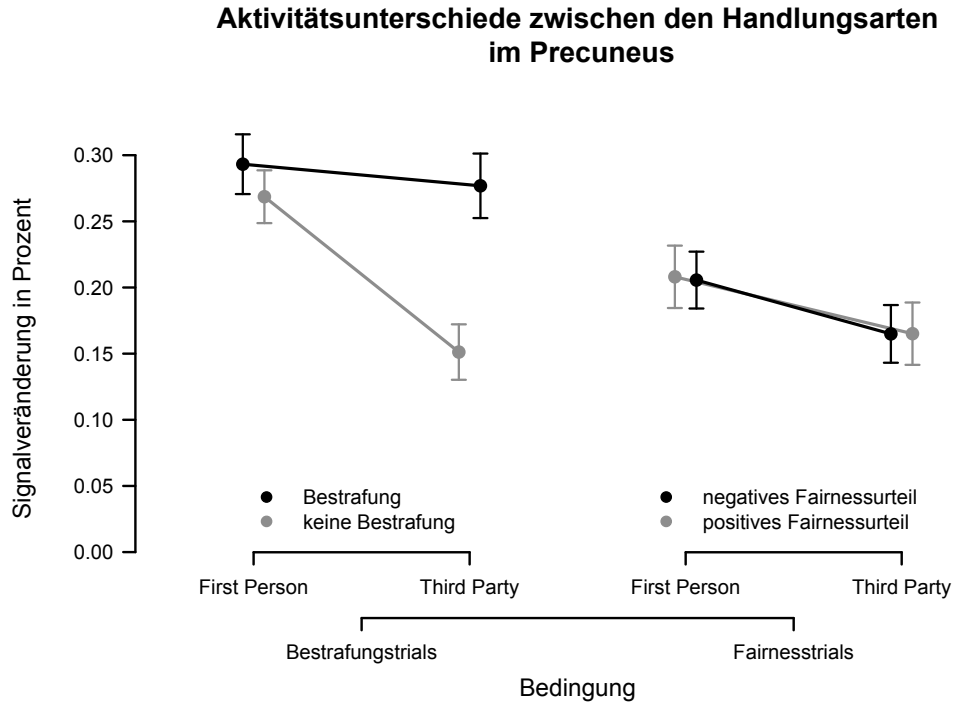

Abbildung 18: Prozentuale Aktivitätsunterschiede im Precuneus aufgeschlüsselt nach Bedingung, Perspektive und Handlungsarten. Fehlerbalken entsprechen einer Standardfehlerabweichung. Tal.-Koordinaten (x, y, z): 13 -59 46.

eine erhöhte Signalveränderung des Precuneus messbar, wenn die Probanden das Dictator Game in der First Person Perspektive spielten. Allerdings zeigte der Trend bei den Bestrafungsdurchgängen davon eine Abweichung, die für den signifikanten Signalunterschied beim Hauptkontrast „Bestrafung versus Fairness“ im Precuneus verantwortlich sein könnte (siehe 3.2.1.3). In Abbildung 18 ist zu sehen, dass sich keine generell höhere Aktivität des Precuneus bei Bestrafungsdurchgängen finden ließ. Dieses Aktivitätsmuster wäre aufgrund der Ergebnisse der Second-Level Analyse zu erwarten gewesen. Stattdessen war das Signal lediglich dann deutlich höher, wenn in der Third Party Bedingung bestraft wurde (im Vergleich zur neuronalen Signalstärke, wenn ein negatives Fairnessurteil in der Third Party Bedingung abgegeben wurde).

Eine vierstufige Varianzanalyse konnte den bereits an den Kontrastbildern (siehe Abbildung 12) erkennbaren Effekt der Perspektive (First Person vs. Third Party) und Bedingung (Bestrafungsdurchgänge vs. Fairnessdurchgänge) bestätigen (Perspektive: $F_{(1,46)} = 4.67$, $p = .03$, $\eta^2 = .02$; Bedingung: $F_{(1,46)} = 15.18$, $p < .001$, $\eta^2 = .06$). Der signifikante Haupteffekt Bedin-

gung (Bestrafungsdurchgänge vs. Fairnessdurchgänge) konnte wie vermutet auf die signifikante Interaktion Perspektive x Bestrafungshandlung zurückgeführt werden ($F_{(1,46)} = 5.12$, $p = .02$, $\eta^2 = .02$).

3.2.2.4 Dorsales Striatum

Auch das dorsale Striatum wurde mit Hilfe der Event-Related Averages näher analysiert. Besonders in dieser Hirnregion, in der Aktivitätsunterschiede in vielen Studien mit ähnlichem Paradigma nachgewiesen wurde (de Quervain et al., 2004; Spitzer, Fischbacher, Herrnberger, Gron & Fehr, 2007), sollte diese Methode zu präziseren Schlussfolgerungen über das Aktivitätsschema führen.

Der in der Second-Level Analyse gezeigte Effekt der Perspektive (Hauptkontrast: „First Person versus Third Party"; siehe 3.2.1.2) zeigte sich deskriptiv stärker in Bestrafungsdurchgängen (siehe Abbildung 19). Die Signaldifferenz zwischen First Person und Third Party war bei Bestrafungsdurchgängen größer als bei Fairnessdurchgängen (Differenz zwischen First Person und Third Party in Bestrafungsdurchgängen: 0.07% gegenüber 0.03% in Fairnessdurchgängen). Diese Effektdifferenz war jedoch nicht statistisch bedeutsam. Eine Varianzanalyse zeigte, dass die Interaktion erster Ordnung, Bedingung (Bestrafungsdurchgang vs. Fairnessdurchgang) x Perspektive (First Person vs. Third Party), nicht wesentlich zur Varianzaufklärung der Signaländerung beitrug ($F_{(1,46)} = 0.7$, $p = .40$, $\eta^2 < .01$).

In Abbildung 19 ist darüber hinaus erkennbar, dass insbesondere bei Bestrafung eine hohe Aktivität im dorsalen Striatum zu verzeichnen war. Varianzanalytisch war der Effekt der Bestrafungshandlung hochsignifikant (Faktor Bestrafungshandlung: $F_{(1,46)} = 10.9$, $p < .001$, $\eta^2 = .05$). Die bedeutsam höhere Aktivität des dorsalen Striatum bei Bestrafung zeigte sich nicht auf Kontrastebene (siehe 3.2.1.1). Eine Erklärung für diese divergierenden Ergebnisse könnte die sehr strenge FDR-Korrektur liefern, die bei der Second-Level Analyse eingesetzt wurde.

Neben dem Effekt des Faktors Bestrafungshandlung ist in Abbildung 19 eine weitere Interaktion zwischen Perspektive (First Person vs. Third Party) und Handlungsart (Bestrafung vs. keine Bestrafung) erkennbar. In der First Person Bedingung war die Aktivitätshöhe des dorsalen Striatum relativ unabhängig davon, ob Strafpunkte vergeben wurden oder nicht. Dagegen zeigte sich in der Third Party Bedingung ein geringer Aktivitätsunterschied im Vergleich zur Ruhebedingung, wenn nicht bestraft wurde und ein zur First Person Perspektive vergleichbar hoher

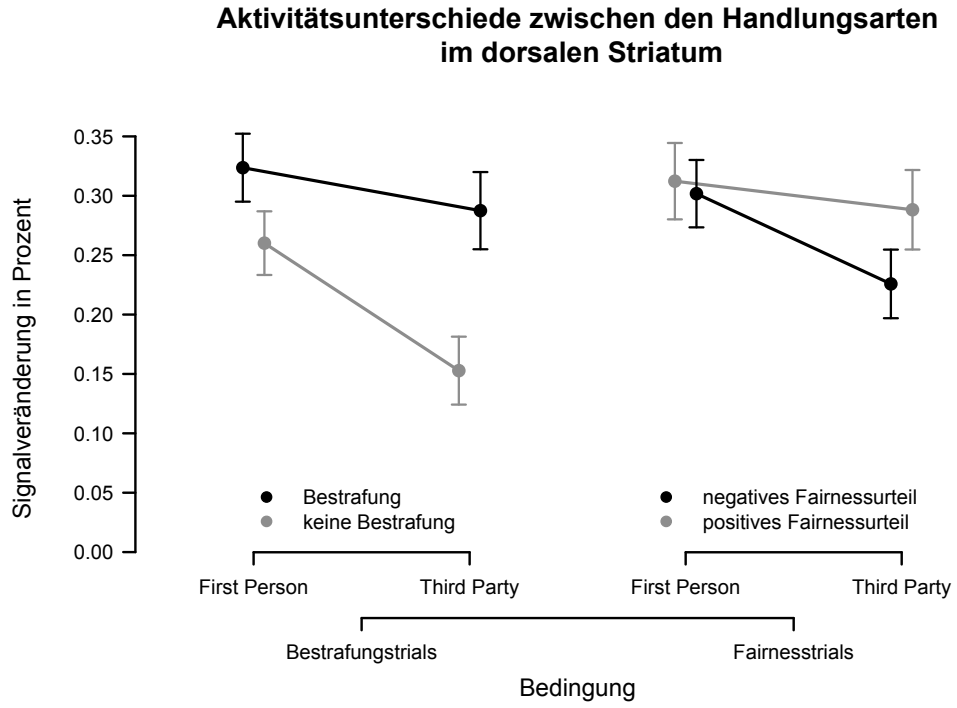

Abbildung 19: Prozentuale Aktivitätsunterschiede im dorsalen Striatum aufgeschlüsselt nach Bedingung, Perspektive und Handlungsarten. Fehlerbalken entsprechen einer Standardfehlerabweichung. Tal.-Koordinaten (x, y, z): -11 10 15.

Aktivitätsunterschied, wenn bestraft wurde. Die deskriptiv erkennbare Interaktion Perspektive x Bestrafungshandlung war allerdings nicht statistisch bedeutsam ($F_{(1,46)} = 1.4$, $p = .23$, $\eta^2 = .01$). Im Gegensatz zu anderen Regionen ist die generell vergleichbar hohe Aktivität in Fairnessdurchgängen im Vergleich zu der in Bestrafungsdurchgängen auffallend. Dies erklärt, warum in der Second-Level Analyse beim Kontrast „Bestrafungsdurchgänge größer Fairnessdurchgänge“ entgegen der Vermutung keine signifikante Aktivitätsdifferenz im dorsalen Striatum festgestellt werden konnte.

3.2.2.5 Anteriorer cingulärer Cortex

Im ACC zeigten sich keine Interaktionseffekte zwischen den Bedingungen und Handlungsstufen. Durch die Methode der Event-Related Averages zeigte sich konsistent mit den Befunden der Second-Level Analyse, dass die relative Aktivität in diesem Hirngebiet dann am größten war, wenn bestraft wurde und auf einem relativ gleichem Niveau lag, wenn nicht bestraft wurde oder Fairnessurteile abgegeben wurden (Abbildung: siehe Anhang C-2).

3.3 Differentialpsychologische Ergebnisse

3.3.1 Ungerechtigkeitssensibilität

In Anlehnung an Schmitt et al. (2005) wurde erwartet, dass Probanden, die sich als sensibel für Ungerechtigkeit beschrieben, wenn sie diese beobachten, bereit sind mehr Punkte zur Bestrafung eines für sie ungerechten Dictator-Verhaltens einzusetzen. Diese Hypothese konnte durch die Daten nicht gestützt werden. Der Wert der zweiten Skala des SBI-Fragebogens korrelierte nicht mit dem Mittelwert der vergebenen Strafpunkte (r = -0.08).

Neuronal sollte eine verzerrte, egozentrische Wahrnehmung von Ungerechtigkeit mit einer höheren Insula-Aktivität bei Bestrafungsverhalten in der First Person Bedingung einhergehen. Dagegen wurde erwartet, dass bei Probanden, die angaben eher Ungerechtigkeit aus einer Beobachterposition heraus wahrzunehmen und weniger gegenüber sich selbst, eine relativ höhere Insula-Aktivität messbar ist bei Bestrafungsverhalten in der Third Party Bedingung.

Um dies zu überprüfen, wurden zunächst für jeden Probanden ein Differenzwert aus der ersten Skala (*Sensibilität für Ungerechtigkeit gegenüber einem selbst*) und der zweiten Skala (*Sensibilität für die Beobachtung von Ungerechtigkeit*) des SBI gebildet. Ein positiver Differenzwert ergab sich demnach, wenn eine Person angab sehr sensibel auf Ungerechtigkeit gegen sich selbst zu reagieren, dagegen jedoch nicht im gleichen Maße Ungerechtigkeit bei anderen wahrzunehmen (und umgekehrt). Ein Differenzwert von null entsprach einer Ungerechtigkeitswahrnehmung, die von der Perspektive unabhängig war. Diese Differenzwerte wurden mit den Differenzwerten der Insula-Aktivität bei Bestrafungsverhalten zwischen First Person und Third Party Bedingung korreliert. Bei der Insula entsprach ein positiver Differenzwert einer relativ zur Third Party Bedingung höheren Aktivität in der First Person Bedingung, wenn bestraft wurde (und umgekehrt).

Der Korrelationskoeffizient zwischen den beiden Differenzwerten war in der linken Insula erwartungsgemäß positiv. Diese Korrelation konnte jedoch statistisch nicht abgesichert werden (r = .27, $t_{(21)}$ = 1.29, p = .21; 95%-Konfidenzintervall: -.16 bis .61). In der rechten Insula zeigte sich kein Zusammenhang im erwarteten Trend (r = 0.03). Probanden mit einer egozentrisch gefärbten Fairnesswahrnehmung zeigten keine höhere Aktivität in der rechten Insula, wenn sie aus der First Person Perspektive bestraften (und umgekehrt).

3.3.2 Empathie

Das Persönlichkeitskonstrukt Empathie wurde mit Hilfe der Skala *Empathiefähigkeit* des SPF und der Skala *Empathie* des I7 Fragebogens erfasst. Die beiden Skalenwerte korrelierten zwischen den Probanden mit r = .46 nur mäßig hoch, was auf die mäßig hohe (SPF) bis geringe (I7) interne Konsistenz der Skalen zurückzuführen ist (α_{SPF} = .70; α_{I7} = .41). Aufgrund der geringen Messgenauigkeit des I7 und der damit verbundenen Schwierigkeit der Interpretation der Messwerte beschränkt sich die folgende Analyse zum Einfluss der Empathie ausschließlich auf die Skala des SPF.

Erwartet wurde eine positive Korrelation zwischen der Ausprägung der Empathiefähigkeit und den vergebenen Strafpunkten in der Third Party Bedingung. Probanden, die angaben hoch empathisch zu sein, sollten sich besser in die Person C hineinversetzen können, dadurch unfaire Aufteilungen stärker wahrnehmen und darauf mit der Vergabe von mehr Strafpunkten reagieren (siehe 1.2.2.3). Neuronal sollte Empathiefähigkeit vor allem bei Bestrafungshandlungen mit einer erhöhten Insula-Aktivität und ACC-Aktivität in der Third Party Bedingung assoziiert sein. Die Hypothese auf Verhaltensebene konnte nicht gestützt werden. Die Empathiefähigkeit korrelierte nicht mit der Anzahl an investierten Strafpunkten in der Third Party Bedingung (r = -.08). Dagegen zeigte sich hypothesenkonform eine erhöhte Insula-Aktivität bei Probanden, die sich als empathisch eingestuft hatten, im Vergleich zu Probanden, die angegeben hatten weniger empathisch zu sein. Der Score im SPF korrelierte mit r = .37 (linke Insula) und r = .54 (rechte Insula) mäßig beziehungsweise mittelhoch. Der Zusammenhang der Einschätzung der Empathiefähigkeit mit der linken Insula war marginal signifikant. Das 95%-Konfidenzintervall der geschätzten Korrelation reichte bis in den negativen Bereich ($t_{(21)}$ = 1.82, p = .08; 95%-Konfidenzintervall: -.05 bis .68). Demgegenüber wich die Korrelation zwischen SPF und rechter Insula hochbedeutsam von null ab (r = .54, $t_{(21)}$ = 2.90, p < .001; 95%-Konfidenzintervall: .16 bis .78). Probanden, die sich als empathischer beschrieben, zeigten demnach ein erhöhte Aktivität in der rechten Insula, wenn sie aus der Beobachterperspektive heraus bestraften.

Es zeigte sich darüber hinaus eine positive, jedoch nicht signifikante Korrelation zwischen dem SPF-Score und der ACC-Aktivität bei Bestrafungshandlungen in der Third Party Bedingung (r = .34, $t_{(21)}$ = 1.67, p = .11; 95%-Konfidenzintervall: -.0.08 bis .66).

3.3.3 Reziprozität

Entgegen der Erwartung hing die Ausprägung in negativer Reziprozität nicht mit dem Bestrafungsausmaß in der First Person Bedingung zusammen. Der Mittelwert der vergebenen Strafpunkte korrelierte nicht (r = -.02) mit dem Testwert des PNRQ. Auch die Aktivitätshöhe im dorsalen Striatum konnte nicht mit dem Grad an berichteter negativer Reziprozität in Zusammenhang gebracht werden. Die Korrelation von r = -.15 war statistisch nicht bedeutsam ($t_{(21)} = 0.71$, p = .49; 95%-Konfidenzintervall: -.53 bis .28).

4 Diskussion

4.1 Ergebniszusammenfassung

Die Ergebnisse zeigen, dass die Probanden sich in der vorliegenden Studie stark negativ reziprok nach der Definition von Fehr et al. (2003; siehe 1.1.3.2 starke Reziprozität) verhielten. Sie opferten eigene Ressourcen in Form von Geldbeträgen, um eine Person für unkooperatives Verhalten zu bestrafen. Sie bestraften demnach altruistisch (Boyd et al., 2003; Fehr & Gächter, 2002). Dies taten sie nicht nur in der Situation, in der das Verhalten der anderen Person einen direkten negativen Einfluss auf sie hatte. Stattdessen nahmen die Versuchspersonen Kosten in Kauf, obwohl das Verhalten der Person, das sie bestraften, in keiner direkten Beziehung zu ihnen selbst stand. Manche Probanden investierten sogar mehr, wenn sie eine unfaire Zuteilung lediglich beobachteten, als wenn sie selbst betroffen waren.

Anhand der Verhaltensdaten konnte gezeigt werden, dass das Ausmaß der Bestrafung zum einen eng an die Fairnesswahrnehmung der Probanden geknüpft war und zum anderen stark von den Aufteilungen der Spieler A abhing. Hierbei lässt sich die Fairness als intermediierende Variable zwischen Angebotshöhe und Bestrafung auffassen. Probanden, die das Verhalten einer Person als unfair einstuften, waren dem zufolge auch bereit diese Person für ihr unfaires Verhalten zu bestrafen. Ob die Probanden direkt oder gar nicht von dem Verhalten, das sie bewerten sollten, betroffen waren, hatte keinen Einfluss auf die Fairnesswahrnehmung.

Des Weiteren konnte nachgewiesen werden, dass Probanden länger für eine Entscheidung brauchten, wenn sie nicht unmittelbar in die Interaktion der Spielsituation involviert waren (Third Party Bedingung). Allerdings konnten keine erhöhten Reaktionszeiten für Bestrafungsverhalten

nachgewiesen werden. Ob Probanden sich entschieden zu bestrafen oder nicht zu bestrafen wirkte sich nicht auf die Reaktionszeit aus.

Auf neuronaler Ebene konnte gezeigt werden, dass der ACC eine erhöhte Aktivität bei Bestrafungsverhalten aufwies, insbesondere, wenn die Probanden direkt in die Spielsituation involviert waren (First Person Bedingung). Dagegen war bei Fairnessurteilen der ACC nicht beziehungsweise sehr gering aktiv. Auch die Insula und der DLPFC wiesen bei Bestrafungsverhalten die höchste Aktivität auf. Sowohl die linke als auch die rechte Insula zeigten eine erhöhte Aktivität in der First Person Bedingung, auch wenn nicht bestraft wurde. In der rechten Insula war die Aktivität sogar annähernd gleich hoch, unabhängig davon, ob sich die Probanden entschieden zu bestrafen oder nicht. Dieser Trend war im DLPFC ähnlich. Auch hier zeigte sich, dass bei keiner Bestrafung eine hohe neuronale Aktivität zu verzeichnen war, sofern die Probanden sich in der First Person Bedingung befanden. Dieser Trend kann als Regression der durchschnittlichen Aktivität hin zur Mitte beschrieben werden. Im Vergleich zu Third Party Durchgängen war in First Person Durchgängen der Unterschied in der Aktivitätshöhe zwischen Bestrafungsverhalten und der Unterlassung einer Bestrafung in diesen Arealen durchweg kleiner. Das dorsale Striatum war sowohl bei Bestrafungsdurchgängen als auch bei Fairnessdurchgängen aktiv. Bei Handlungen in der First Person Bedingung im Vergleich zur Third Party Bedingung war diese Region stärker aktiv. Bei näherer Analyse konnte gezeigt werden, dass das dorsale Striatum auch bei Bestrafungshandlungen im Vergleich zu keiner Bestrafung ein erhöhtes neuronales Signal aufwies. In Bestrafungsdurchgängen und in Durchgängen, in denen Probanden unmittelbar mit Spieler A interagierten, war der PC aktiver. Ein ähnliches Muster zeigte sich im Precuneus. Auch hier hing eine hohe Aktivität mit Handlungen aus der First Person Perspektive sowie mit Bestrafungshandlungen zusammen.

Unerwartet war die erhöhte Aktivität in Regionen des Thalamus bei Bestrafungsdurchgängen und in der First Person Bedingung. Auch über den inferioren parietalen Lobus, der in First Person Durchgängen eine hohe Aktivität aufwies, und den mittleren frontalen Gyrus, der bei Bestrafungsdurchgängen verstärkt aktiv war, wurden a priori keine Hypothesen zur Beteiligung am Phänomen der altruistischen Bestrafung aufgestellt.

Differentialpsychologisch wurde vermutet, dass das Ausmaß an Bestrafung, welches die Probanden in der Beobachterrolle (Third Party) zeigten, sich mit der Ausprägung in den Konstrukten

Empathie und Ungerechtigkeitssensibilität aus einer Beobachterperspektive erklären lässt. Beide Ansätze zur Erklärung von Unterschieden im Bestrafungsverhalten konnten durch die Daten nicht gestützt werden. Auch die Ausprägung in negativer Reziprozität hing nicht wie vermutet mit dem Bestrafungsausmaß in der First Person Bedingung zusammen.

Dagegen konnte gezeigt werden, dass empathische Personen insbesondere in der rechten Insula eine erhöhte Aktivität aufwiesen, wenn sie aus einer Beobachterrolle heraus altruistisch bestraften. Zwar zeigte sich bei diesen Personen auch tendenziell eine erhöhte ACC-Aktivität, jedoch konnte dieses Ergebnis statistisch nicht abgesichert werden.

Weiterhin wurde die Hypothese getestet, ob Personen, die angaben, bei Betrug oder Täuschung mit Rache oder Vergeltung zu reagieren (negative Reziprozität), eine erhöhte Aktivität in Belohnungsarealen, wie dem N. caudatus, aufweisen, wenn sie sich für unfaire Aufteilungen des Spielers A im Dictator Game mit Bestrafungspunkten revanchieren. Auch diese Hypothese konnte durch die Daten nicht gestützt werden. Ferner wurde die These geprüft, nach welcher Personen, die eher Ungerechtigkeit gegenüber sich selbst wahrnehmen, eine höhere Insula-Aktivität in den Bedingungen aufweisen, in denen sie als Spielpartner agieren im Vergleich zu der Situation, in der sie nur als Beobachter handeln konnten. Lediglich für die linke Insula, nicht jedoch für die rechte Insula, konnte ein Trend in dieser Richtung festgestellt werden. Allerdings konnte aufgrund fehlender statistischer Power nicht hinreichend nachgewiesen werden, dass dieser Trend nicht auch durch zufällige Fehlerschwankung in den Daten erklärbar wäre.

4.2 Einbettung der Ergebnisse in die bisherige Befundlage

Die starke Verknüpfung von Fairness und Bestrafung und die Konsistenz im Bestrafungsverhalten über den Perspektivenwechsel (First Person und Third Party) hinweg ist ein Beleg dafür, dass eine Fairnessnorm als Motiv hinter dem Verhalten steht (Fehr & Schmidt, 1999). Auch die Fairnesswahrnehmung war unabhängig von der Perspektive. Dadurch wird die Annahme von einer Norm unterstützt, welche intersubjektiv als Verhaltensrichtlinie und sozialer Wertestandard anerkannt wird und unabhängig der eigenen Interessen gilt (siehe 1.2.2.1).

Das dennoch ein Konflikt zwischen Eigeninteresse und der Durchsetzung einer Fairnessnorm besteht, konnte auf Verhaltensebene durch die verzögerte Reaktion der Probanden festgestellt werden, wenn diese zunächst nicht unmittelbar in das Spiel involviert waren. Dieses Ergebnis kann in Einklang mit dem Befund von Knoch et al. (2006) gebracht werden. Die Autoren konn-

ten analog zu unserem Ergebnis beobachten, dass Personen in einem Ultimatum Game dann längere Reaktionszeiten aufwiesen, wenn sie ungleiche Angebote zurückwiesen und damit ihren eigenen Geldgewinn minimierten (siehe 1.4.4.3). Dass Personen jedoch nicht länger brauchten, um zu entscheiden, ob sie bestrafen wollen oder nicht, könnte ein Hinweis darauf sein, dass insbesondere in der First Person Bedingung die Entscheidung stärker an Motive wie Rache oder Revanche gekoppelt ist und durch diese affektive Komponente im Entscheidungsprozess eine rationale Kosten-Nutzen Rechnung unterdrückt wird. In der Third Party Bedingung, in der keine direkte Verbindung zum Dictator besteht, könnte dagegen die Kosten-Nutzen Kalkulation stärker in den Entscheidungsprozess einfließen.

Die Rolle des N. accumbens als „Motivator" hinter zunächst irrational erscheinendem Bestrafungsverhalten, die vor allem bei Strobel et al. (in prep.), aber auch bei Harbaugh et al. (2007) hervorgehoben wurde, konnte in der vorliegenden Studie nicht gestützt werden. In Gebieten des ventralen Striatums, zu dem der N. accumbens gezählt wird, konnten keine bedeutsamen Aktivitätsunterschiede ausgemacht werden. Auch Gebiete des dorsalen Striatums zeigten nicht das erwartete Aktivitätsmuster. Zwar war eine erhöhte Aktivität des dorsalen Striatums bei Bestrafungshandlungen messbar, was im Einklang mit der Interpretation von de Quervain et al. (2004) steht. Die Forschergruppe postulierte, dass insbesondere der N. caudatus mit der Befriedung assoziiert ist, die vermeintlich entsteht, wenn eine Fairnessnorm durchgesetzt werden kann. In der vorliegenden Studie zeigte sich allerdings eine ähnlich hohe Aktivität des dorsalem Striatum, wenn die Probanden Fairnessurteile abgaben. Man könnte vermuten, dass selbst die Abgabe eines Fairnessurteils zu Befriedigung führt, da damit die eigenen Fairness-Maßstäbe nach außen getragen werden können. Dennoch legt das Ergebnis nahe, dass die Aktivität im dorsalen Striatum nicht den Stellenwert bei dem Phänomen der altruistischen Bestrafung einnimmt, wie die bisherigen Studien vermuten lassen (siehe dazu Sanfey et al., 2006).

Im Einklang mit bisherigen Untersuchung und unseren Hypothesen konnte dagegen die Insula, der ACC sowie der DLPFC als wichtige Regionen des Gehirns für das Phänomen der altruistischen Bestrafung identifiziert werden (Buckholtz et al., 2008; Sanfey et al., 2006; Strobel et al., in prep.). Insbesondere die linke Insula zeigte bei Bestrafungshandlungen eine erhöhte Aktivität. Der PC war besonders in First Person Durchgängen stärker aktiv (vgl. Strobel et al., in prep.).

Unsere Befunde unterstützen die Hypothese, dass die Insula auf einer affektiv-emotionalen Ebene auf Normbruch reagiert und damit einen Erklärungsansatz für den Einsatz von Bestra-

fung liefert (Sanfey et al., 2003). Relativ war die Aktivität in der rechten Insula dann besonders hoch, wenn aus einer Third Party Perspektive heraus bestraft wurde. Eine Erklärung für dieses Ergebnis kann hier das Konzept der Empathie liefern, das von Vignemont und Singer (2006) und de Waal (2007) mit der Insula in Verbindung gebracht wurde. Personen, die sich in der vorliegenden Studie als empathisch beschrieben, zeigten eine relativ höhere Insula-Aktivität, wenn sie subjektiv als unfair wahrgenommene Aufteilungen beobachteten und dieses bestraften. Der vermeintlich entstandene Ärger, an dessen Entstehung die Insula möglicherweise beteiligt ist, wäre aus diesem Erklärungsansatz heraus der Mediator zwischen Empathie und Altruismus, wie dies die Empathie-Altruismus-Hypothese postuliert (Batson & Moran, 1999; siehe 1.2.2.3). Allerdings investierten empathische Personen nicht mehr, zeigten auf Verhaltensebene also nicht mehr altruistische Bestrafungshandlungen, was zunächst einen Widerspruch darstellt. Im Rahmen eines Kosten-Nutzen Ansatzes könnte ein Grund für die Beobachtung, dass sich die erhöhte Aktivität der Insula bei empathischen Personen nicht in einem erhöhten altruistischen Bestrafungsverhaltens manifestierte sein, dass höhere kognitive Kontrollinstanzen, die möglicherweise im ACC verortet sind, hier eine dominantere Rolle bei der Handlungssteuerung besitzen, insbesondere wenn Probanden nur als Beobachter agieren. Hier wären weitere Experimente, in denen stärker auf die Empathie-Altruismus-Hypothese fokussiert wird, notwendig (siehe 4.4). Für die Empathie-Altruismus-Hypothese spricht auch eine aktuelle Studie von Yamagishia, Horita, Takagishi, Shinada, Tanida & Cook (2009), die durch verschiedene Varianten eines Ultimatum-Games zeigen konnten, dass Konzepte wie Ungleichheitsaversion oder Reziprozität nur geringfügig zur Erklärung des altruistischen Bestrafungsverhaltens beitragen konnten. Stattdessen lieferten emotionale Prozesse (wie Ärger oder moralischer Ekel) eine bessere Erklärung für die erhobenen Daten. Dem analog konnte auch in der vorliegenden Studie keine bedeutsame Verknüpfung von Ungerechtigkeitssensibilität und negativer Reziprozität mit dem Bestrafungsverhalten nachgewiesen werden.

Mit Bezug auf die Ergebnisse von Sanfey et al. (2003) und anderen neuroökonomischen Studien (Mcclure, Laibson, Loewenstein & Cohen, 2004; Strobel et al., in prep.) konnte der DLPFC auch in der vorliegenden Studie mit dem Phänomen der altruistischen Bestrafung in Verbindung gebracht werden. Dass diese Hirnregion dabei stark an die Bestrafungshandlung gekoppelt ist, konnte dadurch gezeigt werden, dass bei Fairnessbewertungen keine erhöhte Aktivität im DLPFC zu verzeichnen war. Der DLPFC scheint demnach an der Ausführung von altruistischer

Bestrafung beteiligt zu sein und weniger an der Evaluation von beobachteter Unfairness beziehungsweise der Evaluation von Normbrüchen. Dies konnten auch Knoch et al. (2006) im Rahmen eines TMS-Experiment zeigen. Durch die systematischen Deaktivierung des DLPFC mit Hilfe der TMS-Methode änderte sich das Bestrafungsverhalten, nicht jedoch die Fairnesseinschätzung (siehe 1.4.4.3).

Die erwähnte „Regression zur Mitte“ in den gemessenen neuronalen Signalstärken des DLPFC und der Insula könnte durch eine Konfundierung der Prozesse stattgefunden haben, die hinter dem Verhalten in der First Person Situation stehen. Während in der Third Party Bedingung eine Person durch die Vergabe von keinen Strafpunkten vollkommen unbeteiligt bleiben kann, so ist die Person in der First Person Bedingung immer involviert in die Situation, ob sie bestraft oder nicht. Es wäre daher möglich, dass der Proband, der das Dictator Game aus einer First Person Perspektive spielt, auch wenn er sich dazu entschließt nicht zu bestrafen, ähnliche Entscheidungsprozesse durchläuft, wie in der Situation, in der er bestraft. Dahingegen liegt in der Third Party Bedingung eine Entkopplung der Interaktion von der Person vor. So ist der Prozess, der hinter der Bestrafung steht, distinkter trennbar, weil die Person die Möglichkeit hat Beobachter zu bleiben.

Im ACC zeigte sich ein solcher Trend nicht. Das Aktivitätsmuster im ACC unterstützt die These, dass diese Hirnregion vor allem an der Kosten-Nutzen Abwägung beteiligt ist, die erst in Bestrafungshandlungen in dem gewählten Paradigma einsetzt. Dies steht im Einklang mit Befunden von Botvinick, Cohen & Carter (2004) und Sanfey et al. (2003).

Interessant und nicht erwartet waren die unterschiedlichen Aktivitätsmuster, die sich im Precuneus zeigten. Der Precuneus wird mit selbstbezogenen kognitiven Prozessen assoziiert (Cavanna & Trimble, 2006). Damit übereinstimmend fand sich eine höhere Precuneus-Aktivität, wenn Personen direkt in die Situation involviert waren. Interessant dabei war, dass die Precuneus-Aktivität ein ähnlich hohes Aktivitätsniveau annahm, wenn Personen zwar nicht selbst betroffen waren (Third Party Bedingung), sich jedoch in die Situation einmischten, indem sie altruistisch bestraften. Somit kann auch diese Situation als teilweise selbstbezogen interpretiert werden und würde die Theorie unterstützen, dass der Precuneus eine wichtige Rolle bei kognitiven Prozessen spielt, in deren Zentrum die eigene Perspektive und das eigene Handeln stehen.

Darüber hinaus waren weitere Hirnregionen aktiv, über deren Rolle bei dem Phänomen der altruistischen Bestrafung bisher wenig bis nichts bekannt ist. Darunter befand sich der inferior-

parietale Lobus, der besonders dann aktiv war, wenn Personen selbst von unfairem Verhalten betroffen waren und dieses bestraften sowie Teile des Thalamus (darunter der medial-dorsale Nucleus), der bei Bestrafungshandlungen und bei First Person Durchgängen aktiver war. Die erhöhte Thalamus-Aktivität könnte durch eine unterschiedliche Informationsverarbeitung zwischen verschiedenen Reizsituationen im Experiment hervorgerufen worden sein und wäre damit stark an die Operationalisierung des Experiments gekoppelt und weniger an das Phänomen der altruistischen Bestrafung. Auch der inferior-parietale Lobus, der vor allem eine wichtige Rolle bei somatosensorischen Prozessen spielt (Blakemore & Frith, 2005), scheint eher auf unterschiedliche Anforderungen in der experimentellen Situation zurückführbar zu sein und weniger auf Bestrafungshandlung oder Fairnessevaluationen.

4.3 Ein Entstehungsmodell der altruistischen Bestrafung – Resümee

Seit den 60er Jahren des letzten Jahrhunderts wurden verstärkt verschiedene theoretische Konzepte zur Erklärung von altruistischen Verhaltensweisen in mehreren wissenschaftlichen Disziplinen wie der Biologie, der Ökonomie und der Psychologie entwickelt. Mit Hilfe des experimentellen Designs unserer Studie wurden Versuchspersonen in eine Situation versetzt, in der sie die Option hatten altruistisch zu handeln. Dabei wurden die Rahmenbedingungen für bestimmte Erklärungsansätze von Altruismus systematisch manipuliert und eliminiert.

So interagierten die beteiligten Personen in diesem Experiment vollkommen anonym. Dadurch konnte ausgeschlossen werden, dass Mechanismen der Verwandtschaftsselektion Einfluss auf das Verhalten der Personen hatten (siehe 1.1.3.1). Auch Konzepte der direkten und indirekten Reziprozität, die prosoziales Verhalten im Rahmen einer Tit-for-Tat-Strategie erklären, können aufgrund der hergestellten Anonymität im Experiment das altruistische Bestrafungsverhalten nicht erklären (siehe 1.1.3.2). Als mögliche Erklärung bleibt daher das Konzept einer sozialen Norm, konkret einer Fairnessnorm, dass heißt eine internalisierte Bewertungsinstanz, die es ermöglicht Fairness unabhängig der eigenen Interessen einzuschätzen (siehe 1.2.2.1 und 1.3.2.2). Die Daten unterstützen diese These (siehe 3.1.1). Dennoch kann das recht abstrakte soziologische Konzept einer Norm aus einer psychologischen Perspektive noch nicht hinreichend erklären, warum Personen dazu bereit sind aktiv einzugreifen, um eine solche Norm wiederherzustellen, obwohl diese Wiederherstellung objektiv mit Kosten und keinen direkten Nutzen (z.B. in Form von Reziprozität) verbunden ist. Hier wurden emotionale Prozesse als mögliche Media-

toren angeführt, insbesondere Prozesse, die durch Empathie hervorgerufen werden (siehe 1.2.2.2 und 1.2.2.3).

Im Rahmen einer weiter gefassten Interpretation der Befunde vermuten wir, dass Empathie eine Schlüsselrolle bei der Erklärung von prosozialen Verhaltensweisen einnehmen könnte. Die Theorie, dass sich Menschen in andere hineinversetzen können und, wie es die Simulation-Theory postuliert (Gallese & Goldman, 1998), in der Lage sind den aktuellen (emotionalen) Zustand einer Person durch Beobachtung nachzufühlen, könnte eine Reihe an sozialen Interaktionsformen erklären, die schlussendlich zu scheinbar irrationalem, der Kooperationsbereitschaft in einer Gesellschaft aber sehr zuträglichem Verhalten wie der altruistischen Bestrafung führt. Wir vermuten im Einklang mit anderen Forschergruppen (Batson & Moran, 1999; Vignemont & Singer, 2006; de Waal, 2007) und ausgehend von unseren empirischen Befunden, dass die Insula bei solchen empathischen Prozessen eine wichtige Rolle spielt. Dieses entwicklungsgeschichtlich alte kortikale Areal könnte in der Entwicklungsgeschichte eine neuronale Basis für die Entwicklung von emotional-gefärbten empathischen Impulsen dargestellt haben, die durch die Beobachtung von Artgenossen in kritischen Situation hervorgerufen wurde. Die Möglichkeit, auf Leiden anderer mit Empathie zu reagieren, könnte wiederum für das Aufkommen von prosozialen Verhaltensweisen in sozialen Gruppen verantwortlich sein. In höheren Entwicklungsstufen, die mit dem Ausbau von komplexeren kognitiven Funktionen einhergehen, könnten sich diese zunächst affektiven und automatischen Prozesse in abstrakte Konzepte wie Fairness und Gerechtigkeit umgewandelt haben, die sich schlussendlich in modernen gesellschaftlichen Institutionen wie Justizwesen oder exekutiven Kontrollinstanzen (z.B. das Polizeiwesen) manifestieren. Die altruistische Bestrafung kann in diesem Kontext als Sonderform von Altruismus betrachtet werden, die vor allem der Aufrechterhaltung von Kooperationen sowohl in archaischen wie modernen Gesellschaften dient, indem Free Rider für egoistische Handlungen bestraft werden (siehe 1.2.1 und 1.2.2.1; Fehr & Gächter, 2002)

Es zeigte sich jedoch, dass eine erhöhte Insula-Aktivität nicht mit mehr altruistischer Bestrafung einherging. Hier könnte ein multipler Ansatz zur Entscheidungsfindung die theoretische Grundlage für die empirischen Befunde liefern (siehe 1.4.2). Höhere kognitive Kontrollinstanzen haben in einem neuronalen System womöglich die Funktion der Überwachung von emotionalen Impulsen und einen größeren Einfluss auf das tatsächliches Verhalten, indem sie die impulsive Handlungstendenz einer Kosten-Nutzen Kalkulation unterwerfen. Dies muss nicht zwangs-

läufig zu einer Unterdrückung von altruistischem Verhalten führen. In unserer Studie zeigten neokortikale Gehirnregionen wie der ACC und der DLPFC eine erhöhte Aktivität bei altruistischer Bestrafung. Es wäre möglich, dass in diesen Regionen diese Kosten-Nutzen Kalkulation durchgeführt wird. Mit größerer Sicherheit kann man sagen, dass diese Areale entscheidend am Entscheidungsprozess und damit der altruistischen Bestrafung beteiligt sind. Weitere Untersuchungen werden nötig sein, diesen Erklärungsansatz mit Hilfe empirischer Forschung zu untermauern und die Rolle der einzelnen Hirnareale distinkter bestimmen zu können sowie die Interaktion zwischen den Hirnregionen im Rahmen eines Netzwerkes näher zu beleuchten.

4.4 Grenzen der Studie und Ausblick

In den folgenden Ausführungen soll ein kritischer Blick auf Methodik und Ergebnisinterpretation geworfen werden. Dadurch soll ein konstruktiver Ansatz für zukünftige Studien im Bereich der Altruismusforschung entwickelt werden.

Die in der vorliegenden Studie berichteten Korrelationen zwischen Fragebogendaten, Verhaltensdaten und MRT-Signalintensitäten lagen durchweg im niedrigen bis mittleren Bereich und nur ein Teil der Korrelationen konnten statistisch abgesichert werden. Dass keine höheren Korrelationskoeffizienten detektiert werden konnten, lässt sich auf die geringe Messgenauigkeit der eingesetzten Instrumente zurückführen. So kann der Korrelationskoeffizient zwischen zwei Konstrukten maximal die Höhe der Wurzel aus dem Produkt der Reliabilitäten der Messinstrumente annehmen[1]. Die durch den Koeffizienten von Cronbach geschätzte Messgenauigkeit (Cronbach, 1951) der Fragebogenskalen schwankte zwischen α = .41 und .85. Schätzungen zur Reliabilität von fMRT-Messungen gehen davon aus, dass diese selten Werte über .70 erreichten (Aron, Gluck & Poldrack, 2006; Vul et al., 2009). Demnach lag die obere Grenze für den Korrelationskoeffizienten in der Analyse der differentialpsychologischen Befunde je nach Fragebogenskala zwischen r = .54 und .77. Als statistisch bedeutsam konnten Zusammenhänge bei der vorliegenden Stichprobengröße erst ab r = .42 ausgemacht werden. Bei Berücksichtigung der geschätzten Messgenauigkeit in diesem Experiment ist eine Stichprobe mit der Mindestgröße von N = 30 für zukünftige Studien zu empfehlen.

[1] $r_{a,b\,\text{(gemessen)}} = r_{a,b\,\text{(tatsächlich)}} \cdot \sqrt{\text{Reliabilität}_a \cdot \text{Reliabilität}_b}$ – (Cohen, Cohen, West & Aiken, 2003)

Bei der vorliegenden Untersuchung ist darüber hinaus zu beachten, dass auch auf neuronaler Ebene mit korrelativen Daten gearbeitet wurde. Dementsprechend sind streng kausale Aussagen nicht möglich. Die gefundenen Wirkungszusammenhänge können also nicht gerichtet interpretiert werden. Auch die funktionale Interpretation der einzelnen Hirnregionen im Rahmen eines multiplen Ansatzes ist mit Vorsicht zu betrachten. Hierfür wären weitere Analysen zum Beispiel mit dem Verfahren der *Granger Causality* hilfreich, die allerdings den Rahmen dieser Arbeit gesprengt hätten.

Dennoch konnte gezeigt werden, dass die Aktivitätsdifferenzen in den beteiligten Hirnarealen besonders groß waren, wenn die Probanden sich in einer Third Party Bedingung befanden. Für zukünftige Studien sollte auf diese Interaktionsform im Rahmen von spieltheoretischen Paradigmen besonders Wert gelegt werden, da altruistische Bestrafung hier in einer „reineren“ Ausprägung betrachtet werden kann, weil Personen hier die Möglichkeit haben, völlig unbeteiligt zu bleiben. Dies wirkt sich, so legen es die Daten nahe, nicht auf auf behaviorale, wohl aber auf die neuronaler Ebene aus.

Bisher weitgehend unbeachtet sind mögliche Geschlechtseffekte beim Phänomen der altruistischen Bestrafung. Dies gilt vor allem für den Forschungszweig, in dem Methoden der Neurowissenschaft eingesetzt werden. Auch in der vorliegenden Studie wurden Unterschiede zwischen den Geschlechtern vernachlässigt. Aufgrund der ungleich großen Verteilung in der Stichprobe (19 Frauen und 7 Männer) schienen Analysen in diese Richtung wenig fruchtbar.

Auch die Interpretation der Ergebnisse auf Verhaltensebene kann unter folgendem Aspekt debattiert werden. Wir können nicht ausschließen, dass die Probanden durch den stetigen Wechsel von Fairnessbeurteilungen und Bestrafungshandlungen in einen kognitiven Dissonanzzustand geraten wären, wenn sie nicht gemäß ihres Fairnessurteils bestraft hätten. Unter dieser Annahme müsste man davon ausgehen, dass die hohe Verknüpfung von Fairnessurteil und Bestrafungsausmaß zum Teil durch das Bedürfnis der Probanden nach konsistentem Handeln erklärt werden kann (Festinger, 1978). Da jedoch sowohl die Angebote als auch die Handlungsoptionen in randomisierter Reihenfolge erfolgten, lag in den meisten Fällen eine zeitliche Distanz zwischen der Fairnessbewertung und der Bestrafungshandlung eines gleich hohen Angebots. Dies sollte einen kognitiven Dissonanzeffekt vermieden oder zumindest abgeschwächt haben. Für zukünftige Studien wäre es dennoch empfehlenswert Fairnessbeurteilungen und Bestrafungshandlungen

in separaten Gruppen zu messen. Da dies jedoch die doppelte Probandenanzahl erfordern würde, ist abzuwägen, ob sich der damit verbundene zusätzliche ökonomische Aufwand lohnt.

Ein interessantes Ergebnis der hier besprochenen Untersuchung war die enge Verknüpfung von Empathie und Insula-Aktivität. Es erscheint vielversprechend in zukünftigen Studien den Fokus stärker auf dieses Phänomen zu richten, zumal bisher wenige Untersuchung die konkrete Verbindung von Empathie und Altruismus beziehungsweise altruistischer Bestrafung auf neuronaler Ebene untersucht haben. Vorstellbar wäre ein Experiment, in dem versucht wird das Ausmaß an Empathie kontrolliert zu manipulieren. Im Dictator Game könnte beispielsweise ein weiterer experimenteller Faktor eingeführt werden, der aus zwei Stufen besteht: Zum einen beobachten die Probanden die Interaktion zwischen zwei unbekannten Personen und haben anschließend die Möglichkeit zu bestrafen (wie in der Third Party Bedingung in diesem Experiment), zum anderen befinden sich die Probanden in der selben Situationen, bekommen aber gleichzeitig ein Bild der Person gezeigt, die gerade unfair behandelt wird. Dies würde über eine gesteigerte Identifikationsmöglichkeit mit der Person empathische Prozesse begünstigen. Man könnte noch eine dritte experimentelle Stufe einfügen, in der die Probanden den Spieler, der vom Dictator Aufteilungen diktiert bekommt, vor dem Experiment persönlich kennenlernt. Damit hätte man drei Stufen, auf denen die Möglichkeit für empathiebezogene Prozesse systematisch manipuliert werden würde.

Darüber hinaus wurde im Rahmen eines Stufenmodells argumentiert, dass entwicklungsgeschichtlich alte Regionen durch evolutionär neuere Regionen mithilfe von kognitiven Kontrollmechanismen „überwacht“ werden (siehe 1.4.2). Dieser Mechanismus, so die Theorie, ermöglicht altruistische Verhaltensweisen, die der Interaktion in Gruppen dienlich sind. Darüber hinaus schützen sie das Individuum vor zu großen Kosten oder Gefahren, die bei altruistischen Verhaltensweisen entstehen können. Um die Rolle des ACC und des DLPFC im Rahmen dieses Ansatzes näher zu untersuchen, könnte es vielversprechend sein das in dieser Studie genutzte Dictator Game Paradigma um einen Faktor „Kostenintensität“ zu erweitern. Man könnte systematisch die Kosten für das Verteilen von Strafpunkten variieren und dabei auf Verhaltensebene beobachten, ab wann sich die Versuchspersonen bei gleicher Aufteilung dagegen entscheiden altruistisch zu bestrafen und welche Aktivitätsunterschiede in den angesprochenen Regionen mit diesem Wechsel im Verhalten einhergehen. So ließe sich die postulierten Kosten-Nutzen Prozesse hinter dem altruistischen Verhalten weiter eingrenzen und untersuchen. Auch der Wendepunkt,

ab dem altruistisches Verhalten in egoistisch-gefärbtes Verhalten umschwenkt, ließe sich durch diese weitere Manipulation beobachten.

Erweiterte Versuchsdesigns, spezifische Analysemethoden, die Bezugnahme auf allgemeinpsychologische und differentialpsychologische Aspekte sowie die Integration von Methoden und Modellen verschiedener wissenschaftlicher Disziplinen tragen dazu bei, das Phänomen der altruistische Bestrafung immer tiefgehender zu untersuchen und verstehen zu können. Ein grundlegendes wissenschaftliches Verständnis des Verhaltens bei gesunden Menschen kann dazu beitragen, gravierende Abweichungen hiervon, zum Beispiel im pathologischen oder kriminellen Kontext, besser interpretieren zu können und adäquate Behandlungsmöglichkeiten zu finden.

Doch möchten wir auch darauf verweisen, dass selbst das „beste“ Modell zur Erklärung von altruistischem Verhalten nicht dazu führen sollte, dass dieses Verhalten mit seinen positiven Konsequenzen an Bedeutung verliert. Ob es dem Motiv nach nun als psychologischer Altruismus, als biologischer Altruismus (siehe 1.1.1) oder gar als egoistisch (Fehr et al., 2003) zu bezeichnen ist, wenn eine Person in einer wohltätigen Einrichtung unentgeltlich aushilft, ändert nichts an der Tatsache, dass dieses Verhalten eine Nutzen für andere hat und dem gemeinschaftlichen Zusammenleben hilft.

Literatur

Adams, S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (pp. 267–299). New York: Academic Press.

Allingham, M. (2002). *Choice Theory – A very short introduction*. Oxford: Oxford University Press.

Aron, A., Gluck, M. & Poldrack, R. (2006). Long-term test–retest reliability of functional MRI in a classification learning task. *NeuroImage*, *29*, 1000–1006.

Asendorpf, J. B. (2005). *Psychologie der Persönlichkeit*. Heidelberg: Springer Verlag.

Axelrod, R. & Hamilton, W. (1981). The evolution of cooperation. *Science*, *211*, 1390–1396.

Balleine, B., Delgado, M. & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience*, *27 (31)*, 8161–8165.

Batson, C. & Moran, T. (1999). Empathy-induced altruism in a prisoner's dilemma. *European Journal of Social Psychology*, *29*, 909–924.

Batson, C. D. (1991). *The altruism question: toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.

Batson, C. D. (1998). Self-other merging and the empathy-altruism hypothesis: reply to Neuberg et al. (1997). *Journal of Personality and Social Psychology*, *73*, 517–522.

Batson, C. D., Fultz, J. & Schoenrade, P. A. (1987). Critical self-reflection and self-perceived altruism: when self-reward fails. *Journal of Personality and Social Psychology*, *53*, 594–602.

Blakemore, S.-J. & Frith, U. (2005). *The learning Brain*. Oxford: Blackwell Publishing.

Bolton, G. E. & Ockenfels, A. (2000). A theory of equity, reciprocity, and competition. *The American Economic Review*, *90 (1)*, 166–193.

Boone, J. (1998). The evolution of magnanimity. *Human Nature*, *9 (1)*, 1–21.

Botvinick, M. M., Cohen, J. D. & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, *8 (12)*, 539–546.

Boyd, R., Gintis, H., Bowles, S. & Richerson, P. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, *100 (6)*, 3531–3535.

Boyed, R. & Richerson, P. (2004). *The Nature of Cultures*. Chicago: University Chicago Press.

Boynton, G. M., Engel, S. A., Glover, G. H. & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience*, *16*, 4207–4221.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D. & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, *60 (5)*, 930–940.

Burnstein, E., Crandall, C. S. & Kitayama, S. (1994). Some neo-darwinian decision rules for altruism: weighing cues for inclusive fitness as a function of the biological importance of the decision. *Journal of Personality and Social Psychology*, *67*, 773–789.

Buss, D. M. (2005). *The handbook of evolutionary psychology*. Hoboken, NJ: Wiley.

Buxton, R. B. (2002). *Introduction to functional magnetic resonance imaging: principles and techniques*. Cambridge: Cambridge University Press.

Calder, A., Lawrence, A. & Young, A. (2001). Neuropsychology of fear and loathing. *Nature Reviews Neuroscience*, *2*, 352–363.

Camerer, C. (2003). *Behavioral game theory: experiments in strategic interaction*. New York: Russel Sage Found.

Camerer, C., Loewenstein, G. & Prelec, D. (2005). Neuroeconomics: how neuroscience can inform economics. *Journal of Economic Literature*, *42*, 9–64.

Cavanna, A. E. & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, *129 (3)*, 564–583.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New Jersey: Lawrence Earlbaum Associates.

Cohen, J., Cohen, P., West, S. & Aiken, L. (2003). *Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences* (3rd ed.). New Jersey: Lawrence Earlbaum Associates.

Critchley, H., Elliott, R., Mathias, C. & Dolan, R. (2000). Neural activity relating to generation and representation of galvanic skin conductance responses: a functional magnetic resonance imaging study. *Journal of Neuroscience*, *20 (8)*, 3033–3040.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16 (3)*, 297–334.

Dale, A. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, *8*, 109–114.

Damasio, A., Grabowski, T., Bechara, A. & Damasio, H. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, *3 (10)*, 1049–1056.

Darwin, C. R. (1859). *The origin of species*. London: Murray.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Catalog of Selected Documents in Psychology*, *10*, 85–91.

De Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Buck, A. & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, *305*, 20.

De Waal, F. (2007). Putting the altruism back into altruism: the evolution of empathy. *Annual Reviews*, *59*, 279–300.

Denton, D., Shade, R., Zamarippa, F., Egan, G., Blair-West, J., McKinley, M., Lancaster, J. & Fox, P. (1999). Neuroimaging of genesis and satiation of thirst and an interoceptor-driven theory of origins of primary consciousness. *PNAS*, *96*, 5304–5309.

Derbyshire, S. W., Jones, A. K. & Gyulai, F. (1997). Pain processing during three levels of noxious stimulation produces differential patterns of central activity. *Pain*, *73*, 431–445.

Elster, J. (1989). *The cement of society – a study of social order*. Cambridge: Cambridge University Press.

Elster, J. (1998). Emotions and economic theory. *Journal of Economic Literature*, *36 (1)*, 47–74.

Evans, C. K., Banzett, R. B., McKay, L., Frackowiak, R. S. & Corfield, D. R. (2002). Bold fMRI identifies limbic, paralimbic, and cerebellar activation during air hunger. *Journal of Neurophysiology*, *88*, 1500–1511.

Eysenck, S., Daum, I., Schugens, M. & Diehl, J. (1990). A cross-cultural study of impulsiveness, venturesomeness and empathy: germany and england. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *11*, 209–213.

Falk, A. & Fischbacher, U. (2001). Distributional consequences and intentions in a model of reciprocity. *Annales d'Economie et de Statistique*, *63*, 111–129.

Falk, A. & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*, 293–315.

Fehr, E. & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*, 785–791.

Fehr, E. & Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8 (4)*, 185–190.

Fehr, E. & Fischbacher, U. (2004b). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*, 63–87.

Fehr, E., Fischbacher, U. & Gächter, S. (2003). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, *13 (1)*, 1–25.

Fehr, E. & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114 (3)*, 817–868.

Festinger, L. (1978). *Theorie der kognitiven Dissonanz*. Bern: Huber.

Forsythe, R., Horowitz, J., Savin, N. & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, *6*, 347–369.

Frank, R. (1988). *The strategic role of emotions*. New York: Norton.

Friedman, J. (1971). A non-cooperative equilibrium for supergames. *The Review of Economic Studies*, *38 (1)*, 1–12.

Frijda, N. H. (1988). The laws of emotion. *American Psychologist*, *43 (5)*, 349–358.

Friston, K., Jezzard, P. & Turner, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping*, *1*, 153–171.

Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2 (12)*, 493–501.

Genovese, C., Lazar, N. & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, *15*, 870–878.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, *206*, 169–179.

Gouldner, A. (1960). The norm of reciprocity: a preliminary statement. *American Sociological Review*, *25 (2)*, 161–178.

Hamilton, W. (1964). The genetical evolution of social behaviour I. *Journal of Theoretical Biology*, *7*, 1–16.

Harbaugh, W. T., Mayr, U. & Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, *316*, 1622–1625.

Hein, G. & Singer, T. (2008). I feel how you feel but not always: the empathic brain and its modulation. *Current Opinion in Neurobiology*, *18*, 153–158.

Hill, K. (2003). Altruistic cooperation during foraging by the ache, and the evolved human predisposition to cooperate. *Human Nature*, *13 (1)*, 105–128.

Hintze, J. & Nelson, R. (1998). Violin plots: a box plot-density trace synergism. *American Statistician*, *52*, 181–184.

Hoffman, E., McCabe, K. & Smith, V. (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review*, *86 (3)*, 653–660.

Holler, M. J. & Illing, G. (2003). *Einführung in die Spieltheorie* (5. Aufl.). Berlin: Springer Verlag.

Holmes, A. P. & Friston, K. (1998). Generalisability, random effects & population inference. *NeuroImage*, *7*, 754.

Homans, G. (1961). *Social Behavior: its elementary forms*. New York: Harcourt, Brace & World.

Iadarola, M. J., Berman, K. F., Zeffiro, T. A., Byas-Smith, M. G., Gracely, R. H., Max, M. B. & Bennett, G. J. (1998). Activation of multiple neural networks during acute pain and allodynia evoked by capsaicin assessed with positron emission tomography. *Brain*, *121*, 931–947.

Kable, J. & Glimcher, P. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10 (12)*, 1625–1633.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, *58 (9)*, 697–720.

Kahneman, D. & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, D. R. Davies & J. Beatty (Eds.), *Variants of attention* (pp. 29–61). New York: Academic Press.

Kaplan, H., Hill, K., Lancaster, J. & Hurtado, A. (2000). A theory of human life history evolution: diet, intelligence, and longevity. *Evolutionary Anthropology: Issues*, *9*, 156–185.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*, 829–832.

Levine, D. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, *1*, 593–622.

Logothetis, N., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150–157.

Lotem, A., Fishman, M. & Stone, L. (1999). Evolution of cooperation between individuals. *Nature*, *400*, 226–227.

McAndrew, F. (2002). New evolutionary perspectives on altruism: multilevel-selection and costly-signaling theories. *Current directions in psychological science*, *11 (2)*, 79–82.

Mcclure, S., Laibson, D., Loewenstein, G. & Cohen, J. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503–507.

Milner, B., Petrides, M. & Smith, M. L. (1985). Frontal lobes and the temporal organization of memory. *Human Neurobiology*, *4 (3)*, 137–142.

Nieuwenhuys, R., Voogd, J. & Huijzen, C. (1991). *Das Zentralnervensystem des Menschen*. Berlin: Springer Verlag.

Nowak, M. & Sigmund, K. (1998). The dynamics of indirect reciprocity. *Journal of Theoretical Biology*, *194*, 561–574.

Nowak, M. & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*, 1291–1298.

O'Doherty, J. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology*, *14*, 769–776.

Ogawa, S., Lee, T., Kay, A. & Tank, D. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*, 9868–9872.

Paulus, C. (2009). Der *Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie.* Zugriff am 03.04.2009, von http://psydok.sulb.uni-saarland.de/volltexte/2009/2363/pdf/SPF_Artikel.pdf.

Penner, L., Dovidio, J., Piliavin, J. & Schroeder, D. (2004). Prosocial behavior: multilevel perspectives. *Annual Reviews*, *56*, 366–385.

Perugini, M., Gallucci, M., Presaghi, F. & Ercolani, A. P. (2003). The personal norm of reciprocity. *European Journal of Personality*, *17 (4)*, 251–283.

Petrides, M. & Pandya, D. (2003). The Frontal Cortex. In G. Paxinos & J. Mai (Eds.), *The Human Nervous System* (pp. 951–997). London: Elsevier.

Phillips, M., Young, A., Senior, C. & Brammer, M. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*, 495–498.

Posner, M. & Snyder, C. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S.Dornic (Eds.), *Attention and Performance V* (pp. 669–682). London: Academic Press.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, *83 (5)*, 1281–1302.

Rushworth, M. F. S. (2008). Intention, choice, and the medial frontal cortex. *Annals of the New York Academy of Sciences*, *1124*, 181–207.

Sanfey, A., Loewenstein, G., McClure, S. & Cohen, J. (2006). Neuroeconomics: cross-currents in research on decision-making. *Trends in Cognitive Sciences*, *10 (3)*, 108–116.

Sanfey, A., Rilling, J., Aronson, J., Nystrom, L. & Cohen, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*, 1755–1758.

Schmitt, M. (1996). Individual differences in sensitivity to befallen injustice (SBI). *Personality and Individual Differences*, *21 (1)*, 3–20.

Schmitt, M., Gollwitzer, M., Maes, J. & Arbach, D. (2005). Justice sensitivity: assessment and location in the personality space. *European Journal of Psychological Assessment*, *21 (3)*, 202–211.

Schneider, W. & Shiffrin, R. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological Review*, *84*, 1–66.

Serences, J. (2004). A comparison of methods for characterizing the event-related bold timeseries in rapid fMRI. *NeuroImage*, *21*, 1690–1700.

Shmuel, A., Yacoub, E., Pfeuffer, J., de Moortele, P.-F. V., Adriany, G., Hu, X. & Ugurbil, K. (2002). Sustained negative bold, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. *Neuron*, *36*, 1195–1210.

Sime, J. D. (1983). Affiliative behavior during escape to building exits. *Jounal of Enviromental Psychology*, *3*, 21–41.

Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G. & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, *56*, 185–196.

Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Gallhofer, B., Windmann, S. & Kirsch, P. (in prep.). Altruistic punishment: is it only revenge?

Talairach, J. & Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical.

Tataranni, P. A., Gautier, J. F., Chen, K., Uecker, A., Bandy, D., Salbe, A. D., Pratley, R. E., Lawson, M., Reiman, E. & Ravussin, E. (1999). Neuroanatomical correlates of hunger and satiation in humans using positron emission tomography. *PNAS*, *96*, 4569–4574.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, *20*, 410–433.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46 (1)*, 35–57.

Vignemont, F. D. & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences*, *10 (10)*, 435–441.

Vul, E., Harris, C., Winkielman, P. & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*, 274–290.

Wade, A. R. (2002). The negative bold signal unmasked. *Neuron*, *36 (6)*, 993–995.

Walster, E., Berscheid, E. & Walster, G. W. (1973). New directions in equity research. *Journal of Personality and Social Research*, *25*, 151–173.

Walter, H. (2005). *Funktionelle Bildgebung in Psychiatrie und Psychotherapie*. Stuttgart: Schatthauer GmbH.

Yamagishia, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S. & Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *PNAS*, *106 (28)*, 11520–11523.