Neuronale und psychologische Korrelate sozialer Präferenzen

Kumulative Arbeit

Inaugural-Dissertation
zur Erlangung der Doktorwürde
der
Philosophischen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
zu Bonn

vorgelegt von
Katarina Kuss-Gondorf, geb. Kuss
aus Bonn

Bonn 2018

Gedruckt mit der Genehmigung der Philosophischen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

Zusammensetzung der Prüfungskommission:

Prof. Dr. Michael Wagner (Vorsitzender)

Prof. Dr. Martin Reuter (Betreuer und Gutachter)

PD Dr. Dipl.-Psych. Klaus Fließbach (Gutachter)

Prof. Dr. Ulrich Ettinger (weiteres prüfungsberechtigtes Mitglied)

Tag der mündlichen Prüfung: 14.12.2017

Inhaltsverzeichnis

Abkürzungsverzeichnis 4
Allgemeine Zusammenfassung 5
Überblick über die vorliegende Arbeit 8
1. Prosoziales Verhalten als Gegenstand der Forschung 9
1.1 Psychologie 9
1.2 Ökonomie 11
1.2.1 Ökonomische Verhaltensexperimente als geeignete Paradigmen in den Neurowissenschaften 13
1.3 Neurowissenschaften 15
1.4 Zielsetzung der Studien 19
2. Zusammenfassung der Publikationen 20
2.1 Die SVO-Studie 20
2.2 Die Charity-Studie 27
2.3 Die Effort-Studie 31
3. Diskussion 37
Literaturverzeichnis 46
Danksagung 53
Anhang 54

Abkürzungsverzeichnis

BOLD	blood oxygenation level dependent
CD	costly donation Bedingung
CS	costly social Bedingung
dlPFC	dorsolateraler präfrontaler Cortex
dmPFC	dorsomedialer präfrontaler Cortex
E	Effizienz Bedingung (Englisch: efficiency)
fMRT	funktionelle Magnetresonanztomografie
FWE	Family-wise error
MNI	Montreal Neurological Institut
mOFC	medialer orbitofronaler Cortex
NAcc	Nucelus Accumbens
NCD	non-costly donation Bedingung
NCS	non-costly social Bedingung
PSI	pure self-interest Bedingung
ROI	region of interest
RPE	Reward Prediction Error, Belohnungsvorhersagefehler
SVO	Social Value Orientation
ToM	Theory of Mind
TPJ	temporoparietal Junction
vmPFC	ventromedialer präfrontaler Cortex

Allgemeine Zusammenfassung

Prosoziale Entscheidung sind Entscheidungen, die die Auswirkung auf andere Personen berücksichtigen und von denen andere Personen profitieren. Aktuelle Erkenntnisse der sozialen Neurowissenschaften sowie der Neuroökonomie legen nahe, dass prosoziale Entscheidungen Hirnareale aktivieren, die im Kontext individueller Entscheidungen mit Belohnung assoziiert sind (u.a. ventromedialer prefrontaler Cortex, medialer orbitofrontaler Cortex, Nucleus Accumbens; Ruff & Fehr, 2014). Dies eröffnet die allgemeine Frage, ob prosoziale Handlungen für den Akteur „belohnend" sind und folglich neuronale Indikatoren eines Belohnungsempfindens beobachtet werden können. Darüber hinaus stellt sich die spezifischere Frage, ob soziale und individuelle (die eigene Person betreffende) Entscheidungen von Aktivität in den gleichen Hirnarealen begleitet werden. Zur Beantwortung dieser Forschungsfragen tragen die Erkenntnisse der vorliegenden drei Studien, mit Hilfe der Methode der funktionellen Magnetresonanztomografie (fMRT), bei. Es handelt sich bei den drei Studien um die SVO-Studie, die Charity-Studie und die Effort-Studie. Die SVO- und die Charity-Studie untersuchten neuronale Korrelate prosozialer Entscheidungen in einem ökonomischen Paradigma mit unterschiedlichen Rezipienten (in der SVO-Studie war ein anderer Studienteilnehmer der Rezipient (Kuss et al., 2015), in der Charity-Studie eine Spendenorganisation (Kuss et al., 2013)). Die Effort-Studie (Hernandez Lallement, J.*, Kuss, K.*, Trautner, P., Weber, B., Falk, A., Fliessbach, 2014) erweiterte diese Erkenntnisse um den Aspekt der Leistung bei prosozialen Entscheidungen und ging der Frage nach, ob „verdientes" Geld neuronal anders verarbeitet wird als „geschenktes" Geld.

Sowohl die SVO- als auch die Charity-Studie verwendeten ein Entscheidungsparadigma, welches der Ökonomie angelehnt ist (sog. modifiziertes Diktatorspiel). Dieses ermöglichte die einzelnen Phasen prosozialer Entscheidungen (*vor, während, nach der Entscheidung*) zu untersuchen, wobei der Fokus auf die Belohnungsareale des menschlichen Gehirns gelegt wurde.

In der SVO-Studie (Kuss et al., 2015) fanden sich *während* prosozialer Entscheidungen für eine andere Person Aktivierung in belohnungsassoziierten Arealen *(*ventromedialer prefrontaler Cortex, medialer orbitofrontaler Cortex), sowie Aktivierungen in Arealen, die mit kognitiver Kontrolle und Deliberation assoziiert sind (dorsomedialer prefrontaler Cortex). Somit zeigten sich, neben den belohnungsassoziierten Aktivierungen, neuronale Indikatoren weiterer kognitiver Prozesse, im Sinne der Kontrolle über primär eigensinnige Motive. Zudem liefert die Studie Erkenntnisse bezüglich interindividueller Unterschiede von Probanden mit unterschiedlicher sozialer Wertorientierung (*social value orientation*: prosoziale versus

egoistische Wertorientierung; Van Lange, 1999). Wir fanden behaviorale und neuronale Indikatoren automatisierten prosozialen Verhaltens von prosozialen Probanden, sowie verstärkter deliberativer Prozesse von egoistischen Probanden während prosozialer Entscheidungen. Diese Ergebnisse legen nahe, dass prosoziales Verhalten je nach Ausprägung des Persönlichkeitsmerkmals *social value orientation (SVO)* entweder eher intuitiv ist, oder einer Unterdrückung eigensinniger Impulse und somit kognitiver Ressourcen bedarf.

Im Anschluss an die Entscheidung induzierten wir Belohnungsvorhersagefehler (Englisch: *Reward Prediction Error*, RPE). Belohnungsvorhersagefehler für den eigenen Geldgewinn sind durch neuronale Aktivität im Nucleus accumbens (NAcc) – einem Teil des mediostriatalen Belohnungssystems – repräsentiert (Pagnoni, Zink, Montague, & Berns, 2002; Schultz, 1998). In der Charity-Studie (Kuss et al., 2013) konnten wir durch unsere experimentelle Manipulation erstmals ein äquivalentes RPE-Signal für einen Spenden-Geldbetrag – und somit einen für das materielle Selbstinteresse der Person völlig irrelevanten Geldbetrag – in der gleichen Hirnregion nachweisen. Dies traf nur für Probanden zu, die bereit waren auf den eigenen Gewinn zugunsten der Spendenorganisation zu verzichten und somit auch behavioral demonstrierten, dass sie der Spendenorganisation einen hohen Wert beimessen.

Die Effort-Studie (Hernandez Lallement, J.*, Kuss, K.*, Trautner, P., Weber, B., Falk, A., Fliessbach, 2014) knüpfte an diese Ergebnisse an und erweiterte diese um den Aspekt der Leistungserbringung bei prosozialen Entscheidungen, d.h. konkret inwieweit die Tatsache, ob jemand für einen Geldbetrag eine Leistung erbracht hat, oder nicht (*Windfall-Money*), die neuronalen Reaktionen auf diese Geldbeträge verändert. Die Ergebnisse verdeutlichen, dass das menschliche Gehirn Kontextfaktoren, bzw. die Umstände des Erhalts einer Belohnung, kodiert: Es zeigte sich eine stärkere Assoziation der Aktivierung in Belohnungsarealen (NAcc) mit der Höhe des „verdienten“ Geldes, wenn das Geld durch das Lösen einer anstrengenden Aufgabe geschah. In ähnlicher Weise zeigte sich eine Assoziation mit der Höhe des Verlustes dieses Geldes in der anterioren Insel. Diese Ergebnisse sprechen dafür, dass das Gehirn den *subjektiven* Belohnungswert kodiert und neben dem absoluten Wert einer Belohnung Kontextfaktoren berücksichtigt.

Die Studien zeigen, dass soziale und nicht-soziale Kognition die gleichen Hirnareale aktivieren und ebenso Belohnungszentren des Gehirns während prosozialer Entscheidungen aktiviert sind, wie es während individueller Entscheidungen der Fall ist. Diese Aktivierungen in klassischen Belohnungsarealen des Gehirns können als neuronaler Indikator für den Belohnungswert prosozialen Verhaltens gedeutet werden und legen den verlockenden, jedoch mit Vorsicht zu ziehenden Schluss nahe, dass prosoziales Verhalten belohnend ist. Zudem werden komplexe Kontextinformationen (Umstand des Erhalts einer Belohnung)

durch das menschliche Gehirn kodiert. Dies könnte ein neuronaler Indikator der erhöhten Sensitivität bezüglich Belohnungen und Verlust nach starker Anstrengung sein – ein aus ökologischer Perspektive adaptiver Mechanismus. Zudem bieten die Paradigmen der SVO- und der Charity-Studie mit der Einführung des Belohnungsvorhersagefehlers (RPE) eine geeignete Methode, zwei Ereignisse getrennt zu beobachten, die per se miteinander verknüpft sind: Neuronale Reaktionen in Zusammenhang mit der Belohnung und der Entscheidung, die zur Belohnung führt.

Überblick über die vorliegende Arbeit

Diese Dissertation stellt empirische Arbeiten zum Thema neuronale und psychologische Korrelate sozialer Präferenzen vor. Der Begriff der sozialen Präferenz entstammt den experimentellen Wirtschaftswissenschaften und beschreibt, wie Menschen in ihren Handlungen mögliche Effekte auf andere Menschen berücksichtigen (Fehr & Camerer, 2007). Psychologische Forschung und Theorie thematisieren in erster Linie die Motivation prosozialen Verhaltens (Batson & Shaw, 1991). Die sozialen Neurowissenschaften sowie die Neuroökonomie identifizieren Hirnareale, die während prosozialer Entscheidungen aktiv sind (Fehr & Krajbich, 2014). Im ersten Kapitel dieser Dissertation werden theoretische Hintergründe und Forschungsergebnisse dieser drei Wissenschaften (Psychologie, Ökonomie, Neurowissenschaften) zusammengefasst und Anforderungen an ein experimentelles Paradigma zur Erforschung prosozialen Verhaltens mittels funktioneller Magnetresonanztomografie (fMRT) abgeleitet.

Im zweiten Kapitel der Dissertation werden die Ergebnisse und Schlussfolgerungen der drei Studien zusammengefasst. Alle drei Studien bedienen sich der Methode der funktionellen Magnetresonanztomografie (fMRT).

Das dritte Kapitel diskutiert die Ergebnisse der Studien und stellt Zusammenhänge her. Die Studien tragen zum Erkenntnisgewinn bezüglich neuronaler Grundlagen prosozialer Entscheidungen und deren interindividueller Unterschiede sowie der Verarbeitung relevanter Kontextinformationen im Zusammenhang mit der neuronalen Codierung eines subjektiven Belohnungswertes bei. Ferner weisen sie auf methodische Aspekte der Gestaltung experimenteller Paradigmen der Neuroökonomie hin.

Die Original-Publikationen sind im Anhang zu finden.

1. Prosoziales Verhalten als Gegenstand der Forschung

Prosoziales Verhalten ist Gegenstand unterschiedlichster wissenschaftlicher Disziplinen, wie z.B. der Psychologie, Ökonomie, Biologie und Philosophie (Batson & Powell, 2003; Fehr & Fischbacher, 2003; Hamilton, 1964; Rawls, 1971).

Prosoziales Verhalten ist definiert als ein Handeln, welches das Wohlergehen einer anderen Person erhöht, und mit individuellen Kosten einhergehen kann (Gęsiarz & Crockett, 2015). Biologische respektive ökonomische Theorien gehen davon aus, dass das übergeordnete Ziel jeglichen Verhaltens das Weitergeben eigener Gene, respektive das Maximieren des eigenen Nutzens ist (Fehr & Fischbacher, 2003; Gęsiarz & Crockett, 2015). Diesen Theorien zufolge ist jedes beobachtbare prosoziale Verhalten im Grunde egoistisch motiviert (z.B. durch Reziprozität, Vermeiden von Bestrafung; siehe Kapitel 1.2). Auch in der Psychologie wird die Debatte geführt, ob es „wahren" Altruismus gibt, oder ob prosoziales Verhalten im Grunde egoistisch motiviert ist, wie z.B. durch Reduktion von Erregung (Englisch: *arousal)* (Batson & Shaw, 1991; Cialdini et al., 1987) siehe dazu Kapitel 1.1.

Die ontogenetische Entwicklung prosozialen Verhaltens ist multifaktoriell bedingt: Prosoziales Verhalten hat zum einen biologische Grundlagen und Voraussetzungen (z.B. die Reifung von Hirnarealen) und ist zum anderen Sozialisationseinflüssen unterlegen. Fehr und Kollegen zeigten beispielsweise eine Zunahme prosozialen Verhaltens mit steigendem Alter für Kinder zwischen 4 und 8 Jahren (Fehr, Bernhard, & Rockenbach, 2008). FMRT-Studien legen nahe, dass die Reifung des dorsolateralen präfrontal Cortex (dlPFC) bei der Umsetzung prosozialen Verhaltens relevant ist und bieten damit eine mögliche Erklärung für die Zunahme prosozialen Verhaltens im Laufe der kindlichen Entwicklung (Steinbeis, Bernhardt, & Singer, 2012). Ältere Forschungsarbeiten der Psychologie untersuchten den Einfluss der Umwelt und der Sozialisation (Bandura, 1977). Prosoziales Verhalten wird diesen zufolge im Verlauf der Entwicklung durch unterschiedliche Belohnungen gelernt und gefördert: zunächst durch materielle Belohnungen, dann soziale Belohnungen (Erwartungen und Lob durch Bezugspersonen; Normen) und abschließend durch intrinsische Selbstbelohnungen beim Erfüllen von internalisierten prosozialen Werten (Bandura, 1977).

1.1 Psychologie

Prosoziales Verhalten ist seit langer Zeit Gegenstand psychologischer Forschung. Neben Prozessmodellen, welche Bedingungen beschreiben, unter denen prosoziales Verhalten auftritt (Darley & Latane, 1968; Latane & Darley, 1968; Latane & Nida, 1981), befasst sich

ein umfassender Forschungszweig mit der Motivation prosozialen Verhaltens (z.B. Batson & Shaw, 1991; Cialdini et al., 1987).

Prosoziales Verhalten umfasst eine Bandbreite von Verhaltensweisen, die darauf abzielen eine oder mehrere Personen zu begünstigen (Batson & Powell, 2003), z.B. helfen, kooperieren, spenden. Abzugrenzen von dem Verhalten ist die Frage nach der Motivation, die diesem Verhalten zugrunde liegt. Es wird kontrovers diskutiert, ob prosoziales Verhalten altruistisch oder egoistisch motiviert ist (Batson & Shaw, 1991; Cialdini et al., 1987). In diesem Zusammenhang stellt sich die Frage nach dem ultimativen Ziel des Verhaltens: Wessen Wohlergehen ist das ultimative Ziel des Handelnden?

Altruistische Motivation liegt vor, wenn das Wohlergehen der anderen Person das ultimative Ziel ist. Eine egoistische Motivation liegt vor, wenn das Wohlergehen der anderen Person ein instrumentelles Mittel ist, um das ultimative Ziel des eigenen Wohlergehens zu erreichen (siehe Beispiele im nächsten Absatz). Prinzipiell kann prosoziales Verhalten sowohl altruistisch als auch egoistisch motiviert sein (Batson & Shaw, 1991).

Dies verdeutlicht auch die Empathie-Altruismus-Hypothese (Batson & Shaw, 1991), die ursprünglich eine altruistische Motivation vorsieht. Demnach erhöht Empathie für einen Leidenden die altruistische Motivation das Leiden zu beenden und das Wohlergehen des Leidenden als ultimatives Ziel zu erhöhen. Jedoch ergeben sich auch einige eigennützige Vorteile, die mit prosozialem Verhalten einhergehen und somit mögliche egoistische Alternativerklärungen bieten: Mit Empathie geht auch eine gewisses Erregung und negative Emotion einher (Cialdini et al., 1987). Durch das prosoziale Verhalten reduziert der Handelnde diese negativen Empfindungen. Somit kann ebenso eine eigennützige Motivation der Erregungsreduktion zugrunde liegen. Das Vermeiden von Bestrafung (z.B. extern in Form von negativer sozialer Bewertung durch andere oder intern durch eigene Gefühle von Schuld) und das Erwarten von Belohnung (z.B. extern in Form von positiver sozialer Bewertung durch andere oder intern durch Bestätigung des eigenen Ichs als gute Person) können ebenfalls eigennützige Motivationen prosozialen Verhaltens darstellen.

Eine langjährige Forschungstradition um Daniel Batson und Kollegen kommt zu dem Schluss, dass wahre altruistische Motivation existiert und somit das Steigern des Wohlergehens einer anderen Person das ultimatives Ziel des Handelnden ist (z.B. Batson et al., 1988; Dovidio, Allen, & Schroeder, 1990). Batson und Kollegen schließen dabei nicht aus, dass gleichzeitig eigennützige Vorteile als unbeabsichtigte Konsequenzen des altruistisch motivierten Verhaltens auftreten können (z.B. *arousal* reduzieren; positive Bewertung). Cialdini und Kollegen hingegen, sehen eigennützige Motivationen, die eigene Erregung und negative Emotionen zu regulieren (*negative state relief model*; Cialdini et al., 1987), oder positive Verstärkung zu erhalten (sei es extern, z.B. durch soziale Anerkennung, oder intern, z.B. Bestätigung eines positiven Selbstbildes) als ultimatives Ziel an und somit

prosoziales Verhalten als durch eigennützige Ziele motiviert an (Cialdini et al., 1987; Maner et al., 2002).

Zu dieser Forschungsfrage möchte die vorliegende Arbeit keinen Beitrag leisten. Vielmehr gilt es festzuhalten, dass prosoziales Verhalten sowohl egoistisch als auch altruistisch motiviert sein kann. In beiden Fällen ist anzunehmen, dass prosoziales Verhalten belohnend ist. Denn, neben externen Belohnungen, wie z.B. sozialer Anerkennung, kann das Erreichen von Zielen und das Umsetzen von moralischen Werten (in diesem Falle z.B. der Wert, für das Wohlergehen einer andere Person zu sorgen, oder das Ziel eine „gute", soziale Person zu sein) ebenfalls belohnend sein (Gęsiarz & Crockett, 2015). Die vorliegende Arbeit möchte untersuchen, welche Hirnareale beteiligt sind, wenn prosoziale Entscheidungen getroffen werden, wobei der Fokus auf belohnungsverarbeitenden Hirnarealen liegt.

1.2 Ökonomie

Die klassischen Modelle der Ökonomie betonen Selbstinteresse als Hauptmotiv menschlichen Handelns (z.B. Camerer, 2003; Fehr & Krajbich, 2014). Diesen Modellen liegt das Menschenbild des homo oeconomicus zugrunde, welcher als ein rationaler, ausschließlich den eigenen Nutzen betrachtender und maximierender Akteur beschrieben wird. Es wird angenommen, dass prosoziales Verhalten durch strategisches Selbstinteresse in Form von Reziprozität, Gewinnen sozialer Anerkennung oder Vermeiden von Bestrafung motiviert ist (Fehr & Krajbich, 2014; Gęsiarz & Crockett, 2015). In diesen Fällen handelt eine Person prosozial, nur weil dies in der Folge zu einem eigenen höheren Gewinn führt.

Prosoziales Verhalten kann dadurch motiviert sein, selbst einen guten Ruf zu etablieren und soziale Anerkennung zu erlangen (Fehr & Krajbich, 2014). In Studien wurde gezeigt, dass Probanden sich prosozialer verhielten, wenn ihre Entscheidungen öffentlich gemacht wurden (Bereczkei, Birkas, & Kerekes, 2010) und unbeteiligte Dritte prosoziales Verhalten von Probanden monetär belohnten (Wedekind & Braithwaite, 2002). Anonyme Entscheidungen hingegen gingen mit reduzierter Prosozialität einher (Bereczkei et al., 2010). Diese Befunde verdeutlichen, dass prosoziales Verhalten durch eigensinnige Reputationsmotive motiviert sein kann.

Eine weitere, im Grunde egoistische Motivation prosozialen Verhaltens ist Reziprozität. Reziprozität ist in Umgebungen zu beobachten, die durch wiederholte Interaktionen gekennzeichnet sind. Reziprozität beschreibt das Motiv wohlwollendes (bzw. schädliches) Verhalten einer Person mit eigenem wohlwollendem (bzw. schadendem) Verhalten zu erwidern (in dem Sinne „wie du mir, so ich dir"). Durch Reziprozität motiviertes prosoziales Verhalten wird gezeigt, weil es die Wahrscheinlichkeit erhöht, dass das Gegenüber ebenfalls wohlwollend agiert (Gęsiarz & Crockett, 2015; Rand & Nowak, 2013).

Ein weiteres, strategisches und egoistisches Motiv prosozial zu handeln ist das Vermeiden von Bestrafung. In ökonomischen Spielen werden prosoziale Verhaltensweisen gezeigt, um Bestrafungen in Form von monetären Verlusten zu verhindern. Ein klassisches Paradigma, in dem dies beobachtet werden kann, ist das Ultimatumspiel (Camerer, 2003; siehe auch Kapitel 1.2.1). Hier ist ein hohes monetäres Angebot (im Sinne einer Aufteilung eines Geldbetrags zwischen Angebotssteller und Rezipient) aus Angst vor Ablehnung eines kleinen Angebots durch den Rezipienten wahrscheinlich.

Einige Eigenschaften experimenteller Paradigmen begünstigen das Auftreten prosozialen Verhaltens, welches durch strategisches Selbstinteresse motiviert ist, wie z.B. wiederholte Interaktionen derselben Spielpartner, der Umstand, dass Auszahlungen von mehr als einem Spielpartner abhängig sind, keine Anonymität (siehe Kapitel 1.2.1).

Psychologische sowie ökonomische Forschung verdeutlicht, dass Menschen von den Vorhersagen der klassischen ökonomischen Modelle abweichen und prosoziales Verhalten auch in Abwesenheit strategischen Selbstinteresses zeigen, wie z.B. im Diktatorspiel (Van Lange, 1999; Camerer, 2003). Neuere Entwicklungen der Verhaltensökonomie versuchen durch Theorien sozialer Präferenzen diesem Umstand Rechnung zu tragen (z.B. Charness & Rabin, 2002; Falk & Fischbacher, 2006; Fehr & Schmidt, 1999).

Zentrales Konzept ökonomischer Modelle ist die Nutzenfunktion (Englisch: *utility*; Fliessbach, 2011). Menschen sind bestrebt, die Nutzenfunktion zu maximieren, meist - ausgehend vom homo oeconomicus - im Sinne der Maximierung eigener Auszahlungen (Camerer, 2003). Theorien sozialer Präferenzen gehen davon aus, dass Menschen nicht nur die Auswirkungen auf den eigenen Gewinn berücksichtigen, sondern auch das Ergebnis der Entscheidung für andere Personen (Fliessbach, 2011). Übertragen auf das Konstrukt der Nutzenfunktion bedeutet das, dass die Nutzenfunktion nicht nur durch die eigene Auszahlung beeinflusst wird, sondern auch nicht-monetäre Aspekte wie Fairness, Gleichheit, Effizienz, die Auszahlung der anderen Person etc. berücksichtigt werden (z.B. Fehr & Schmidt, 1999). Personen mit sozialen Präferenzen wägen folglich soziale und egoistische Aspekte gegeneinander ab und integrieren diese in eine Nutzenfunktion (Fehr & Krajbich, 2014). In diesem Sinne steigern Menschen den Nutzen (*utility*), wenn sie sich nach sozialen Normen verhalten (wie z.B. Fairness; Fehr & Schmidt, 1999).

Als weitere Beispiele für Theorien sozialer Präferenzen seien Spendenmodelle angeführt. Vorab sei erwähnt, dass diese wirtschaftswissenschaftlichen Modelle in erster Linie in der Lage sind, Spendenmärkte vorherzusagen. Unter Spendenmarkt ist die Bereitstellung öffentlicher Güter zu verstehen (Konow, 2010). Eine Theorie individueller Spendenmotive, deren Entstehung und Verbindung zu anderen Motiven, ist nicht primär Gegenstand der wirtschaftswissenschaftlichen Modelle und Forschung.

Das Modell des reinen Altruismus (*pure altruism*, Andreoni, 1990) geht davon aus, dass Menschen eine Präferenz für öffentliche Güter haben: Je größer das Angebot eines öffentlichen Gutes ist, desto größer ist der Nutzen (Konow, 2010). Dabei spielt es keine Rolle, von wem das öffentliche Gut zur Verfügung gestellt wird (z.B. durch (eigene) Spenden, Steuern). Demnach sollte der Beitrag individueller Spenden abnehmen, wenn ein öffentliches Gut, beispielsweise durch öffentliche Mittel, finanziert wird (sog. *crowding-out*; Konow, 2010). Diese Theorien gehen von einem primär *ergebnisorientierten* Nutzen aus: der Nutzen steigt mit der Höhe der Spende (unabhängig von der Quelle der Spende) und der Fokus liegt somit auf dem Ergebnis für den Rezipienten. In diesem Zusammenhang sprechen wir in der Charity-Studie von ergebnisorientierten Motiven. Demgegenüber stehen *handlungsorientierte* Motive: Die Theorie des *warm glow of giving* (Andreoni, 1990; Harbaugh, 1998) sieht vor, dass ein Individuum Nutzen aus dem reinen Akt des Gebens zieht. Demnach führt der Akt des Gebens zu einem angenehmen Gefühl. Hier liegt der Fokus auf der Handlung des Gebens, weswegen wir von *handlungsorientierten Motiven* sprechen. Das Model des unvollkommenen Altruismus (*impure altruism*; Andreoni, 1990) kombiniert beide oben beschriebenen Modelle des reinen Altruismus und des *warm glow of giving*. Demnach ist der Nutzen abhängig von der Verfügbarkeit des öffentlichen Guts und wird durch die eigene Spende weiter gesteigert.

1.2.1 Ökonomische Verhaltensexperimente als geeignete Paradigmen in den Neurowissenschaften

Ökonomische Spiele der Spieltheorie (Camerer, 2003) bieten Paradigmen, um soziale Präferenzen zu erforschen, z.B. das Diktatorspiel, Ultimatumspiel, Vertrauenspiel, Gefangenendilemma. In diesen Paradigmen treffen Personen Entscheidungen, die Auswirkungen auf die Auszahlung der Person selbst sowie auf Auszahlungen einer oder mehrerer anderer Personen haben. Eine Person zeigt soziale Präferenzen, wenn sie (einen Teil) ihrer Auszahlung aufgibt bzw. einsetzt, um die Auszahlung einer anderen Person zu beeinflussen; als prosozial wird die Präferenz beschrieben, wenn dadurch die Auszahlung der anderen Person erhöht wird. Die unterschiedlichen Eigenschaften der Spiele ermöglichen das Erforschen unterschiedlicher sozialer Verhaltensweisen und Dynamiken (z.B. Vertrauen, Reziprozität, Kooperation). Die Spiele eignen sich aufgrund ihrer einfachen und klaren Durchführung gut im neurokognitiven Kontext zur Untersuchung neuronaler Grundlagen von Entscheidungsprozessen (Fliessbach, 2011).

Ein einfaches Spiel, in dem soziale Präferenzen frei von strategischem Selbstinteresse gezeigt werden, ist das sogenannte Diktatorspiel (Camerer, 2003). Hier erhält eine Person (der Diktator) einen Geldbetrag (X). Diesen Geldbetrag kann der Diktator zwischen sich und

einem anderen Spieler (Rezipienten) aufteilen. Der Rezipient hat keinen Einfluss auf die Aufteilung des Geldes, es findet keine Interaktion statt. In diesem Spiel geben Personen in der Regel zwischen 0 und 50% des Geldbetrags X ab (Hoffmann et al., 1996). Die Höhe der Abgabe wird unter anderem beeinflusst von Kontextfaktoren (Fehr & Krajbich, 2014), aber auch interindividuellen Unterschieden (Declerck, Boone, & Emonds, 2013; Van Lange, 1999): prosoziale Menschen geben für gewöhnlich mehr ab.

In den vorliegenden Studien wird eine Abwandlung des Diktatorspiels angewendet, ein sogenanntes modifiziertes Diktatorspiel (Charness & Rabin, 2002; Fliessbach, 2011): Hier entscheidet sich eine Person zwischen zwei Alternativen, jede Alternative besteht aus einem Geldbetrag für sie selbst (A) und einen weiteren Geldbetrag für eine andere Partei (B). Die Probanden entscheiden sich folglich zwischen zwei Alternativen: A1 B1 oder A2 B2 (genauere Beschreibung des Paradigmas in Kapitel 2.1 sowie in den Publikationen).

Das Ultimatumspiel stellt eine Erweiterung des Diktatorspiels dar, insofern, dass der Rezipient eine Entscheidung trifft, nämlich das Angebot des ersten Spielers (Angebotssteller) anzunehmen oder abzulehnen. Nimmt der Rezipient das Angebot an, wird das Geld wie angeboten aufgeteilt. Lehnt er ab, gehen beide Personen leer aus. Hier findet eine Interaktion im Sinne der Abhängigkeit der Auszahlung von beiden Spielern statt. Somit sind bei diesen Entscheidungen strategisch-egoistische Abwägungen relevant. Üblicherweise liegen die Angebote im Ultimatumspiel höher als im Diktatorspiel (Camerer, 2003). Das häufigste Angebot im Ultimatumspiel ist die Gleichaufteilung des Geldes, im Diktatorspiel hingegen ist das Null-Angebot das häufigste Angebot (Forsythe et al., 1994).

Wichtige Eigenschaften dieser Spiele zur Erfassung sozialer Präferenzen, die möglichst frei von strategischen, egoistischen Überlegungen sind, sind folgende: es werden Entscheidungen bezüglich realer Geldbeträge getroffen, die Entscheidungen betreffen in der Regel anonyme Partner, die Spiele beinhalten keine Täuschung und werden als sogenanntes *one-Sho*t-Spiel realisiert (Fehr & Krajbich, 2014).

One-Shot bedeutet in diesem Zusammenhang, dass eine Versuchsperson eine einmalige Entscheidung trifft, die dann umgesetzt wird, sowie dass Spielpartner nur einmal miteinander interagieren. Dies zielt darauf ab, für Reziprozität zu kontrollieren, die durch mehrmaliges Interagieren wahrscheinlich wird. Die Methode der funktionalen Kernspintomografie erfordert jedoch das mehrmalige Wiederholen eines Ereignisses. Daher werden fMRT-Experimente, die sich spieltheoretischer Paradigmen bedienen, häufig als „Quasi-one-Shot“-Experiment konzipiert. Hierbei treffen die Probanden eine Vielzahl von Entscheidungen, von denen am Ende des Experiments eine Entscheidung zufällig ausgewählt und umgesetzt wird (Fliessbach, 2011).

Anonymität ist ein weiteres zentrales Merkmal. Anonymität schließt das Motiv der Reputation aus, welches wiederum ein strategisches egoistisches Motiv prosozialen Verhaltens darstellen kann.

Zudem sollte die Entscheidung eines Individuum unabhängig von den Erwartungen bezüglich des Verhaltens der anderen Person sein, z.B. kann ein hohes Angebot im Ultimatumspiel nicht als Indikator einer prosozialen Präferenz gewertet werden, weil dieses Angebot auch von der Erwartung der Handlung der anderen Person beeinflusst wird. Prinzipiell können zwei Motive hinter einem hohen Angebot eines Proposers im Ultimatiumspiel stecken: zum einen das echte Interesse für das Gegenüber („reine soziale Präferenz“), zum anderen die strategische Überlegung, dass ein kleines Angebot abgelehnt werden wird und somit das Angebot, inklusive des eigenen Gewinns, nicht realisiert wird (Fehr & Krajbich, 2014; Fliessbach, 2011). Somit ist das Ultimatumspiel ungeeignet zur Erfassung „reiner“ sozialer Präferenzen, die frei von strategischen, egoistischen Überlegungen sind

Ein weiteres Merkmal der ökonomischen Verhaltensexperimente ist, dass Probanden Entscheidungen bezüglich realer Geldbeträge treffen und sich die Entscheidungen im Experiment auf die Vergütung für die Teilnahme auswirken, somit also auszahlungsrelevant sind. Zudem ist ein weiteres sehr streng durchgeführtes Charakteristikum, die Probanden nicht zu täuschen. In psychologischen Experimenten trifft dies meist nicht zu: oft werden hypothetische Entscheidungen ohne Konsequenzen getroffen.

Gegenstand der vorliegenden empirischen Arbeiten sind neuronale Korrelate sozialer Präferenzen und somit gilt es ein geeignetes Paradigma zu wählen, welches in der Lage ist, „reine“ soziale Präferenzen zu messen, die möglichst frei von strategischem Selbstinteresse sind. Aufgrund dessen nutzten die vorliegenden Studien eine abgewandelte Form des Diktatorspiels (siehe Kapitel 2.1; eine ausführliche Beschreibung des Paradigmas ist den Publikationen zu entnehmen). Zudem wurden die Experimente nach ökonomischen Standards durchgeführt: die Entscheidungen waren auszahlungsrelevant und die Studienteilnehmer wurden nicht getäuscht. In den vorliegenden Studien wurde das Geld tatsächlich an die Rezipienten (Spendenorganisationen bzw. die anderer Studienteilnehmer) überweisen. Zudem erhielten die Probanden eine Auszahlung, die abhängig von ihrem Verhalten in dem Experiment war.

1.3 Neurowissenschaften

Eine interdisziplinäre Sichtweise aus kognitiven Neurowissenschaften, Psychologie und Ökonomie ist die Neuroökonmie (Fehr & Camerer, 2007). Diese konzipiert das Finden einer Entscheidung als einen mehrstufigen Prozess und identifiziert Hirnareale, die typischerweise

während dieser unterschiedlichen Entscheidungsphasen aktiviert sind (Rangel & Hare, 2010).

Ein klassisches Forschungsfeld der (Neuro-) Ökonomie ist Entscheidungsfindung in individuellen, nicht-sozialen Kontexten, wie beispielsweise Kaufentscheidungen. In diesem Zusammenhang findet die Theorie der wertbasierten Entscheidungen (Englisch: *value-based decision making*; Rangel, Camerer, & Montague, 2008) Anwendung. Diese Theorie beschreibt, wie Menschen Entscheidungen treffen und erforscht auf Basis klar definierter Entscheidungskomponenten deren neuronalen Korrelate. In diesem Zusammenhang wurden Hirnareale identifiziert, die ein sogenanntes „Valuation-System" konstituieren, u.a. medialer orbitofrontaler Cortex (mOFC), ventromedialer prefrontaler Cortex (vmPFC), Nucleus Accumbens (NAcc), Insula (Rangel et al., 2008). Es werden drei Arten von subjektiven Werten definiert, die im Verlauf eines Entscheidungsprozesses gebildet werden: Entscheidungswert (*decision value*), Erfahrungswert (*experienced value*) und Vorhersagefehler (*prediction error*). Der Entscheidungswert ist konzipiert als subjektiver, erwarteter Wert einer Wahlmöglichkeit. In diesen Entscheidungswert werden Eigenschaften einer Option, u.a. Kosten und Nutzen integriert und auf dessen Grundlage eine Entscheidung getroffen. Dieser Wert stimmt am ehesten mit dem ökonomischen Konzept der Nutzenfunktion überein (Fehr & Krajbich, 2014). Der Erfahrungswert (*experienced value*) entspricht dem tatsächlichen, erfahrenen subjektiven Wert beim Erhalt der Option. Der Vorhersagefehler beschreibt die Differenz zwischen dem Entscheidungswert und dem Erfahrungswert. Dieser Vorhersagefehler ist essentiell, um einen Lernprozess zu initiieren. Der Vorhersagefehler korrigiert die Wahrnehmung und fördert das Lernen des Organismus bezüglich zukünftigem Verhalten in dem Sinne: war die Belohnung so gut, wie erwartet? Sollte ich mich beim nächsten Mal wieder für diese Alternative entscheiden (Fehr & Krajbich, 2014; Ruff & Fehr, 2014)?

Für Entscheidung im individuellen Kontext (z.B. monetäre Entscheidungen) werden für diese unterschiedlichen Werte konsistent Aktivitäten in den Hirnarealen des „Valuation-Systems" gefunden: im vmPFC werden erwartete Kosten und Nutzen einer Option abgewogen und in den Entscheidungswert integriert (Kable & Glimcher, 2009). Der Erhalt einer positiv bewerteten Option im Sinne einer Belohnung (Erfahrungswert) geht mit Aktivierungen im mOFC einher (Bartra, McGuire, & Kable, 2013). Vorhersagefehler werden in dopaminergen Neuronen des Mittelhirns, v.a. ventalen tegmentalen Region und Nucleus NAcc kodiert (Pagnoni et al., 2002; Schultz, 1998). Diese Areale stellen die Kernareale des Valuation-Netzwerks dar. Sie erhalten Input aus diversen anderen Arealen (z.B. dlPFC, Amygdala, Insula), um u.a. Kontextfaktoren in die Wertberechnung mit einzubeziehen (z.B. Emotionen; Risiko; Zeitpräferenzen) (Hare, O'Doherty, Camerer, Schultz, & Rangel, 2008; Kable & Glimcher, 2007, 2009; Plassmann, O'Doherty, & Rangel, 2007).

Es wird angenommen, dass diese neuronalen Korrelate der Wertberechnung bei jeglicher zielgerichteter Entscheidung zu finden sind. Dies wurde über diverse Entscheidungsdomänen hinweg für unterschiedliche individuelle Belohnungsarten bestätigt, z.B. Geld (Kahnt, Heinzle, Park, & Haynes, 2010; B Knutson, Adams, Fong, & Hommer, 2001; Yacubian, 2006), sexuell attraktive Reize (Bray & O'Doherty, 2007; Kampe, Frith, Dolan, & Frith, 2001), Nahrung (O'Doherty, Deichmann, Critchley, & Dolan, 2002; Plassmann et al., 2007), Getränke (Paulus & Frank, 2003) und Autos (Erk, Spitzer, Wunderlich, Galley, & Walter, 2002). Zudem sind die Aktivierungen für individuell präferierte Objekte höher. Beispielsweise zeigte sich eine stärkere Aktivierung des mOFC für präferierte Süßigkeiten (Plassmann et al., 2007), Kaufprodukte (Knutson, Rick, Wimmer, Prelec, & Loewenstein, 2007), sowie für Präferenzen bezüglich des zeitlichen Aufschubs monetärer Belohnungen (Kable & Glimcher, 2007). In einer Studie von Levy und Kollegen (2011), wurden spätere Entscheidungen für individuell präferierte Medienprodukte (Filme, Bücher, CDs), durch neuronale Aktivierung in mOFC und NAcc vorhergesagt (Levy, Lazzaro, Rutledge, & Glimcher, 2011).

Vor dem Hintergrund, dass prosoziale Entscheidungen intrinsische sowie extrinsische Belohnungsaspekte haben (siehe Kapitel 1; z.B. Konow, 2010), stellt sich die Frage, ob die gleichen Hirnareale des Valuation-Netzwerks, die bei individuellen belohnenden Entscheidungen relevant sind, auch bei Entscheidungen im sozialen Kontext relevant sind.

Hierzu lassen sich zwei Positionen unterscheiden: die *common-currency* Hypothese, und die *social valuation specific* Hypothese (Ruff & Fehr, 2014). Üblicherweise untersuchten neurokognitive Studien im Bereich der sozialen Neurowissenschaften vor allem soziale Kognition, wie Erkennen von Emotionen, *Theory of Mind (ToM)*, Empathie. In diesem Zusammenhang wurden Hirnareale identifiziert, die spezifisch während dieser sozialen Kognition aktiv waren, u.a. dorsomedialer präfrontaler Cortex (dmPFC), superior temporal sulcus, temporal-parietal junction (TPJ) (Adolphs, 2003, 2009). In Anlehnung an diese Befunde nimmt die *social valuation specific* Hypothese an, dass soziale Belohnungen und Werte in dafür spezialisierten Hirnarealen bzw. Neuronenpopulationen verarbeitet werden. Diese Hirnregionen haben sich u.a. evolutionär spezifisch für diese Aufgaben entwickelt (Ruff & Fehr, 2014).

Die *common currency* Hypothese (Ruff & Fehr, 2014) hingegen nimmt an, dass sowohl soziale als auch nicht-soziale Entscheidungen in denselben neuronalen Netzwerken des Valuation-Netzwerks verarbeitet werden. Es wird also angenommen, dass eine domänenübergreifende Wert-Kodierung in demselben Valuation-Netzwerk stattfindet (vmPFC, mOFC, NAcc). Dieser Hypothese zufolge ist es möglich, dass das *common currency* Netzwerk, je nach Entscheidung (sozial / nicht-sozial), Input aus domänenspezifischen Arealen erhält. Im Fall von sozialen Entscheidungen können dies

durchaus die Areale der *social brain* Hypothese sein. Dies ist in einzelnen Studien durch Konnektivitätsanalysen bereits bestätigt: Hare und Kollegen konnten zeigen, dass während Spendenentscheidungen eine stärkere Konnektivität zwischen TPJ und vmPFC besteht (Hare, Camerer, Knoepfle, & Rangel, 2010). Dies legt den Schluss nahe, dass TPJ, als ein typisches Areal der sozialen Kognition, die Wert-Bildung für Spenden im vmPFC beeinflusst.
Zahlreiche fMRT-Studien zeigen Aktivierungen in den klassischen Arealen des individuellen Valuation-Netzwerkes während Entscheidungen und Belohnungsverarbeitung im sozialen Kontext und sind somit im Einklang mit der *common currency* Hypothese (für einen zusammenfassenden Überblick siehe Ruff & Fehr, 2014). Beispielsweise ging positive Bewertung durch andere mit Aktivierung im NAcc einher (Izuma, Saito, & Sadato, 2008). Monetäre Entscheidungen, die andere begünstigen, gingen ebenfalls mit Aktivierungen im Valuation- und Belohnungs-System einher, beispielsweise Spenden für einen guten Zweck (Harbaugh, Mayr, & Burghart, 2007); genauere Diskussion siehe (Kuss et al., 2013), aber auch gegenseitige Kooperation in einem interaktiven Spiel (Decety, Jackson, Sommerville, Chaminade, & Meltzoff, 2004; Rilling et al., 2002). Eine Studie von Tricomi und Kollegen fand Aktivierung im Valuation-System, welche mit dem Modell der Ungleichheitsaversion übereinstimmt: es zeigten sich Aktivierungen in NAcc und vmPFC während Probanden ungleichheitsreduzierende Transfers beobachteten, vor allem wenn der Rezipient „arm“ war (bedingt durch eine vorherige experimentelle Manipulation) (Tricomi, Rangel, Camerer, & O’Doherty, 2010). Dies verdeutlicht zum einen, dass Belohnungen (im Sinne von monetären Transfers) für andere in den gleichen Arealen verarbeitet werden, wie eigennützige Belohnungen. Zum anderen wird deutlich, dass komplexe Kontextinformationen, wie hier das Ausgangsgehalt des Rezipienten, mitkodiert werden (Tricomi et al., 2010). Zaki und Mitchell (2011) fanden neuronale Korrelate im mOFC während prosozialer, effizienter Entscheidungen und deuten diese als Hinweise für das Vorhandensein von intrinsischen sozialen Belohnungswerten für prosoziales Handeln (Zaki & Mitchell, 2011).
Zusammenfassend entsprechen die empirischen Befunde der Vorstellung der *common currency* Hypothese (für weitere Studien siehe zusammenfassend Fehr & Ruff, 2014): Denn prosoziale Entscheidungen, die in Einklang mit sozialen Werten und Prinzipien stehen und andere Menschen begünstigen, gehen einher mit neuronaler Aktivierung in Arealen des klassischen Valuation-Netzwerks, welche ebenfalls mit eigennützigen Belohnungsentscheidungen assoziiert sind (Bartra et al., 2013). Dies spricht für eine domänenübergreifende, neuronale Wertberechnung, die Entscheidungen in diversen Kontexten zugrunde liegt.

1.4 Zielsetzung der Studien

Sowohl ökonomische als auch psychologische Theorien und Studien legen nahe, dass prosoziales Verhalten mit unterschiedlichen Belohnungen einhergehen kann (z.B. Steigerung des Wohlergehens einer anderen Person als ultimatives Ziel, aber auch positive soziale Bewertung, Reduktion negativ erlebter Erregung durch Leid). Es ist folglich anzunehmen, dass ähnliche neuronale Belohnungskorrelate zu beobachten sind, wie bei eigennützigen Entscheidungen (siehe *common currency* Hypothese, Kapitel 1.3).

Ökonomische Paradigmen der Spieltheorie bieten einen geeigneten formalen Rahmen, prosoziale Entscheidungen mittels bildgebender Verfahren zu untersuchen. Im Diktatorspiel treffen Personen Entscheidungen über die Verteilung von Geldbeträgen zwischen sich selbst und einem Rezipienten (z.B. eine andere Person oder eine Spendenorganisation). Dies ermöglicht eine direkte und einfache Operationalisierung prosozialen Verhaltens (siehe Kapitel 1.2.1). Auf Basis der Entscheidungsphasen des *value-based decision making* (Entscheidungswert und Belohnungsvorhersagefehler, siehe Kapitel 1.3) gestalteten wir ein Paradigma des modifizierten Diktatorspiels, um neuronale Korrelate dieser einzelnen Entscheidungskomponenten in sozialen Situationen zu untersuchen.

Im Rahmen der vorgelegten Arbeit soll ein Beitrag zu der offenen Forschungsfrage geleistet werden, ob – im Sinne der *common currency* Hypothese - soziale Entscheidungen mit ähnlichen (Belohnungs-) Aktivierungen einhergehen, wie eigennützige Entscheidungen. Dazu wurden zwei Studien realisiert (sog. Charity-Studie und SVO-Studie), in denen prosoziale Entscheidungen für zwei unterschiedliche Rezipienten (Spendenorganisation in der Charity-Studie; anderer Studienteilnehmer in der SVO-Studie) getroffen wurden. Zudem wurde die Frage nach interindividuellen Unterschieden adressiert. Die Charity-Studie untersuchte, ob Probanden, die bereit waren eigene Gewinne aufzugeben, um der wohltätigen Organisation Geld zu spenden (sog. *Donator*) stärkere neuronale Belohnungskorrelate zeigten als Probanden, die nicht bereit waren, eigene Vorteile aufzugeben (sog. *Non-Donator*). Die SVO-Studie griff zusätzlich die Frage auf, ob prosoziales Verhalten intuitiv ist, oder ob prosoziales Verhalten eine Unterdrückung eigensinniger Motive bedarf. Darüber hinaus wurde der Einfluss des Persönlichkeitsmerkmals der sozialen Wertorientierung (*social value orientation, SVO;* Van Lange, 1999) in diesem Zusammenhang untersucht. Eine dritte Studie, die Effort-Studie, untersuchte, inwiefern das menschliche Gehirn komplexe Kontextinformationen neuronal verarbeitet: Es wurde untersucht, welche Rolle eigene Leistungen bei prosozialen Entscheidungen spielen, d.h. inwieweit die Tatsache, ob jemand für einen Geldbetrag eine Leistung erbracht hat oder nicht (*Windfall-Money*), die neuronalen Reaktionen auf diese Geldbeträge verändert.

2. Zusammenfassung der Publikationen

Das folgende Kapitel fasst die drei Studien der Dissertation zusammen. Alle Studien erfassen mithilfe der funktionellen Magnetresonanztomografie neuronale Korrelate prosozialer Präferenzen.

2.1 Die SVO-Studie

Viele unserer Entscheidungen haben Konsequenzen für unsere Mitmenschen. Somit werden sowohl das eigene Interesse als auch die Konsequenzen für andere in unseren Entscheidungen berücksichtigt und gegeneinander abgewogen. Oft entscheiden sich Menschen prosozial, d.h. sie geben teilweise sogar eigene Vorteile auf, um das Ergebnis für eine andere Person positiv zu beeinflussen (Bogaert, Boone, & Declerck, 2008; Declerck et al., 2013; Fehr & Fischbacher, 2003). Es ist eine offene Debatte in der Neuroökonomie, ob prosoziales Verhalten intuitiv ist, oder das Ergebnis deliberativer Prozesse (Fehr & Camerer, 2007). Aktuelle Studien legen nahe, dass Persönlichkeitsmerkmale (z.B. soziale Wertorientierung, Englisch: social value orientation, SVO; Van Lange, 1999) in diesem Zusammenhang eine Rolle spielen: Es gibt Hinweise dafür, dass Personen mit einer prosozialen Wertorientierung (sog. Prosocials) intuitiv prosozial handeln, während Personen mit einer egoistischen Wertorientierung (sog. Proselfs) ebenfalls prosozial handeln können, dies jedoch einer Unterdrückung egoistischer Impulse bedarf (Bogaert et al., 2008; Declerck et al., 2013).

Zahlreiche Studien zeigen interindividuelle Unterschiede prosozialen Verhaltens (Boone, Declerck, & Kiyonari, 2010; Declerck & Bogaert, 2008; Declerck et al., 2013) sowie assoziierter neuronaler Korrelate zwischen Prosocials und Proselfs (Emonds, Declerck, Boone, Vandervliet, & Parizel, 2011; Haruno & Frith, 2010; van den Bos & Guroglu, 2009; Haruno et al., 2014). Personen mit einer prosozialen Wertorientierung, sog. Prosocials, (Van Lange, 1999), haben eine Präferenz, die Auszahlungen für sich und andere zu maximieren. Eine proself Wertorientierung äußert sich in der Präferenz, lediglich eigene Auszahlungen zu maximieren (Van Lange, 1999). Die Studien zeigen, dass prosoziales Verhalten der Prosocials Eigenschaften automatisierten, intuitiven Verhaltens aufweisen: z.B. gaben Prosoicals mehr Geld an den Rezipienten in einem Diktatorspiel unter kognitiver Belastung (im Folgenden wird der englische Begriff *cognitive load* verwendet*)* als Proselfs (Cornelissen, Dewitte, & Warlop, 2011). Zudem zeigten Prosocials stärkere Reaktionen subkorticaler Strukturen (u.a. Striatum) während der Konfrontation mit prosozialen monetären Aufteilungen (Haruno & Frith, 2010; van den Bos & Guroglu, 2009) (Haruno et al 2014). Proselfs hingegen

handelten intuitiv egoistisch und benötigten kognitive Ressourcen, um egoistische Impulse zu unterdrücken und prosozial zu handeln: Unter *cognitive load* entschieden Proslefs zugunsten ihres eigenen Vorteils (Cornelissen et al., 2011) und zeigten während sozialer Entscheidungen in einem Gefangenen-Dilemma stärkere Aktivierung des dlPFC (Emonds et al., 2011).

Daher wurde dieser Untersuchung die Annahme zugrunde gelegt, dass prosoziale Probanden intuitiv prosozial handeln, proself Probanden hingegen kognitive Ressourcen benötigen, um egoistische Impulse zu unterdrücken und prosozial zu handeln. Die Studie nutzte behaviorale und neuronale Korrelate, um dies während Entscheidungen in einem modifizierten Diktatorspiel zu überprüfen. Das Paradigma ähnelt dem SVO-Maß (Van Lange, 1999) und ist in der Lage, den dem SVO-Konstrukt zugrundeliegenden Entscheidungsprozess näher zu beleuchten.

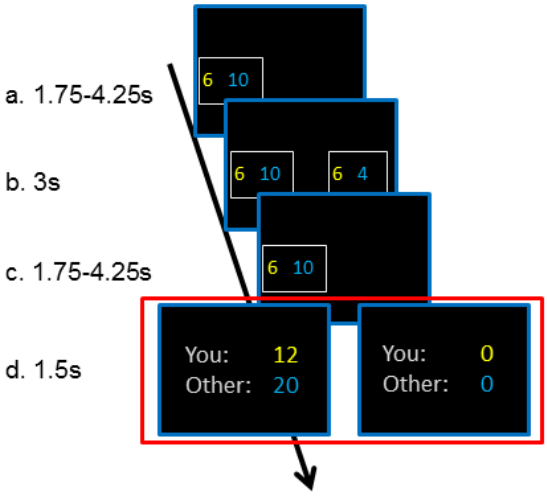

Abbildung 1: Ein Trial des modifizierten Diktatorspiels[1]: Zunächst sehen die Probanden die 1. Alternative (a), bestehend aus einem Geldbetrag für den Diktator (in gelb) und einen Geldbetrag für den Rezipienten (in blau). Die 2. Alternative erscheint und die Probanden wählen eine Alternative (b). Die gewählte Alternative bleibt zu sehen (c). Die Beträge der gewählten Alternative werden entweder verdoppelt, oder auf null gesetzt mit einer 50% Wahrscheinlichkeit (d: Zeitpunkt der RPE-Induktion). Die Geldbeträge sind so gestaltet, dass sich 4 Bedingungen unterschiedlicher Konflikthaftigkeit zwischen egoistischen und prosozialen Motiven ergeben (siehe dazu Tabelle 1).

[1]Das Paradigma der Charity-Studie ist äquivalent. Der Rezipient ist eine selbst gewählte Spendenorganisation. Die RPE-Induktion ist zweistufig: Neben der RPE-Induktion direkt im Anschluss an die Entscheidung (wie in der Abbildung zu sehen), fand eine weitere, zeitlich spätere RPE-Induktion statt, die von der Entscheidung losgelöst war (Details siehe Veröffentlichung)

Die Probanden füllten vor der Teilnahme an dem fMRT-Experiment online den SVO-Fragebogen (Van Lange, 1999) aus und wurden als Proself (n=20) und Prosocial (n=20) klassifiziert.

Im Rahmen der Studie wurde die Bewertung und Integration egoistischer und prosozialer Motive in dem Paradigma des modifizierten Diktatorspiels untersucht (siehe Abbildung 1) mit besonderem Fokus auf die Bedingung, in der nahezu alle Probanden prosozial handeln: die *non-costly social* Bedingung (NCS). Die experimentellen Bedingungen zeichnen sich durch unterschiedliche Konflikthaftigkeit egoistischer und prosozialer Motive aus (Tabelle 1 zeigt einen Überblick der Bedingungen; für Details siehe Publikation): Die *costly social* Bedingung (CS) beinhaltet einen Konflikt zwischen prosozialen und egoistischen Motiven. Hier müssen Probanden eigene Gewinne aufgeben, um prosozial zu handeln. Die Rate prosozialer Entscheidungen ist folglich relativ gering. Die übrigen Bedingungen zeichnen sich durch eine Konfliktlosigkeit aus. Während in der *pure self-interest* Bedingung (PSI), eine Alternative eindeutig vorteilhaft hinsichtlich egoistischer Motive ist, ist in der *non-costly social* Bedingung (NCS) eine Alternative eindeutig vorteilhaft in Bezug auf prosoziale Motive. In dieser Situation können Probanden prosozial handeln (einer anderen Person mehr Geld zukommen lassen), ohne die persönliche Auszahlungen zu beeinflussen. In dieser Bedingung ist auch von den Proselfs eine hohe Rate prosozialer Entscheidungen zu erwarten. Benötigen die Proselfs, im Vergleich zu den Prosocials, kognitive Ressourcen um dies zu tun? Von besonderem Interesse ist folglich der Vergleich der beiden Gruppen während dieser prosozialen Entscheidungen in der *non-costly social* Bedingung.

Hinsichtlich der Ergebnisse gilt es zu berichten, dass sich behavioral eine stärkere Bereitschaft der Prosocials zeigte, eigene Gewinne aufzugeben, um der anderen Person eine höhere Auszahlung zukommen zu lassen: In der *costly social* Bedingung wählten die Prosocials signifikant häufiger die prosoziale Alternative als die Proselfs (siehe Tabelle 1). Dies validiert die Einteilung der Gruppen und bestätigt, dass die Gruppe der Prosoicals trotz eigener Kosten prosozial handelt. In den anderen Bedingungen zeigten sich keine Gruppenunterschiede hinsichtlich des Entscheidungsverhaltens.

Tabelle 1: Häufigkeiten der Wahl der linken Alternative in den vier Bedingungen und Gruppenunterschiede zwischen Prosocials und Proselfs

Bedingung	*Prosocials* (Mittelwert ± SD)[b]	*Proselfs* (Mittelwert ± SD)[b]	*t-Wert (p)*
Pure self-interest (PSI)[a] z.B. **10/6** 4/6	94.1% (± 12.92%)	95% (± 12.11%)	-0.22 (0.829)
Efficiency (E)[a] z.B. **16/10** 4/6	95.5% (± 8.96%)	95% (± 12.53%)	0.14 (0.886)
Non-costly social (NCS)[a] z.B. **6/10** 6/4	90.3% (± 16.25%)	92.1% (± 12.7%)	-0.36 (0.724)
Costly social (CS)[a] z.B. **4/10** 10/6	19.6% (± 16.32%)	6.9% (± 10.94%)	2.79 (0.008)

[a] Die vier Bedingungen[2] mit einem Beispiel einer Entscheidungssituation
[b] Häufigkeiten, die linke der beiden Alternativen zu wählen, gemittelt über alle Trials einer Bedingung getrennt für die Gruppen der Prosocials und Proselfs

Die Analyse der Reaktionszeiten zeigte Unterschiede zwischen den Bedingungen ($F_{(3, 102)}$ = 41.13, $p < 0.001$) sowie eine signifikante Interaktion (Bedingung x Gruppe: $F_{(3, 102)}$ = 14.59, $p < 0.001$).

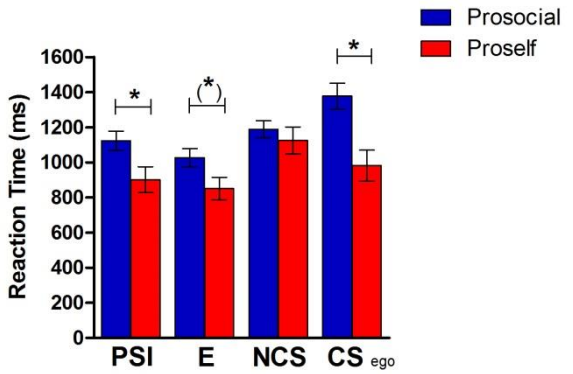

Abbildung 2: Reaktionszeiten in den vier Bedingungen für die Gruppe der Prosocials und Proselfs. PSI: Wahl der eigennützigen Alternative in der *pure-self interest* Bedingung; E: Wahl der effizienten Alternative in der *efficiency* Bedingung; NCS: Wahl der sozialen Alternative in der *non-costly social* Bedingung; CS ego: Wahl der eigennützigen Alternative in der *costly social* Bedingung. *p>0.05; (*)p<0.1.

Prosocials zeigten längeren Reaktionszeiten in allen Bedingungen, wobei die Unterschiede zwischen den Gruppen in den Bedingungen CS und PSI signifikant waren und einen Trend in der E-Bedingung zu verzeichnen war (siehe Abbildung 2). In der NCS-Bedingung zeigte sich kein signifikanter Unterschied. Bemerkenswert ist, dass Prosocials am längsten für

[2] Die Bedingungen in der Charity-Studie sind äquivalent, lediglich die Namen einzelner Bedingungen unterscheiden sich: *Die non-costly social Bedingung* heißt in der Charity-Studie *non-costly donation* Bedingung (NCD); die *costly social* Bedingung heißt *costly donation* Bedingung (CD).

egoistische Entscheidungen in der konflikthaften *costly-social* Bedingung brauchten, während egoistische Probanden die längsten Reaktionszeiten in der konfliktfreien NCS-Bedingung zeigten.

Die NCS-Bedingung stellte die einzige Bedingung dar, in der eine Maximierung des eigenen Gewinns nicht möglich war. Hier waren die Probanden angehalten den Betrag der anderen Person in ihre Entscheidung mit einfließen zu lassen. Dadurch induzierten wir mehr oder weniger prosoziales Verhalten. In dieser Situation waren proself Individuen angeregt, ihre egoistische Grundeinstellung (Englisch: *default)* (Wahl der Alternative mit dem höheren eigenen Betrag) zu überwinden und den Betrag für die andere Person in ihrer Entscheidung zu berücksichtigen. Prosocials hingegen scheinen den Betrag der anderen Person in allen ihren Entscheidungen zu berücksichtigen, was sich in längeren generellen Reaktionszeiten der Prosocials wiederspiegelte.

Des Weiteren stellt sich die Frage, welche neuronalen Aktivierungen während dieser konfliktfreien sozialen Entscheidungen zu beobachten sind. Der Fokus der fMRT-Ergebnisse lag auf dem Vergleich der Bedingungen NCS > PSI. Dieser Vergleich wurde gewählt, weil die Bedingungen konzeptionell identisch waren und sich lediglich hinsichtlich ihres Effekts (in der NCS-Bedingung konfliktfrei prosozial, in der PSI-Bedingung konfliktfrei egoistisch zu handeln) unterschieden. Zudem wurde wie oben ausgeführt in der NCS-Bedingung eine ausreichend hohe Rate prosozialer Entscheidungen beobachtet. In der gesamten Stichprobe zeigte sich eine Aktivierung des vmPFC (NCS > PSI: MNI-Koordinaten des peak Voxels: X = 0,Y = 35,Z = 4, t = 5.53, k = 340, pFWE(whole brain cluster level) < 0.05) sowie des dorsomedialen PFC (NCS > PSI: MNI-Koordinaten des peak Voxels:: X = -9,Y = 56, Z = 34, t = 4.15,k = 204, pFWE(whole brain cluster level) < 0.05). (Abbildung 3). Somit waren während sozialer Entscheidungen Areale aktiv, die einerseits mit subjektiven (Belohnungs-) Wert und Abwägen von Belohnungseigenschaften (vmPFC) sowie andererseits mit Theory of Mind (ToM) und kognitiver Kontrolle (dmPFC), in Zusammenhang gebracht wurden (Rangel and Hare, 2010; Bartra et al., 2013; Saptute and Lieberman, 2006; Lieberman, 2007; Elliot and Dolan, 1998; Gallagher and Frith, 2003; Saxe, 2006).

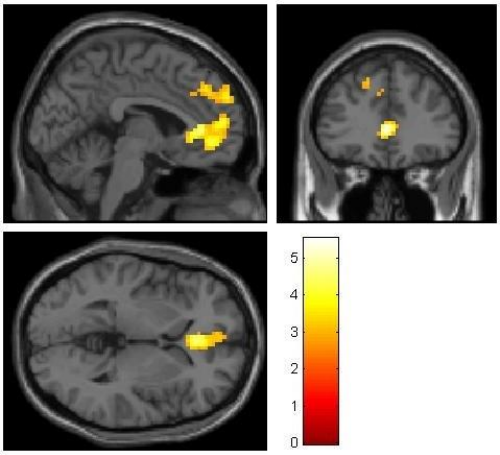

Abbildung 3: Stärkeres BOLD-Signal für soziale Entscheidungen in der NCS-Bedingung im Vergleich zu eigennützigen Entscheidungen in der PSI-Bedingung im vmPFC und dmPFC in der gesamten Stichprobe (n=36). MNI: X = -3, Y = 35, Z = 2, Schwellenwert bei $t > 2.73$, entspricht $p < 0.005$, unkorrigiert.

Im Gruppenvergleich (Proself versus Prosocial) zeigte sich eine stärkere Aktivierung des dmPFC (MNI-Koordinaten des peak Voxels: X = 0,Y = 32,Z = 34, t = 3.86, k = 172, pFWE(whole brain cluster level)<0.05) sowie des mOFC (MNI-Koordinaten des peak Voxels: X = 6,Y = 47,Z = -14, t = 5.01, pFWE(small-volume corrected) < 0.05) in der Gruppe der Proselfs (Proself > Prosocial, Abbildung 4). Die Gruppe der Prosocials zeigte keine stärkeren Aktivierungen im Vergleich zu den Proselfs (Prosocial > Proself).

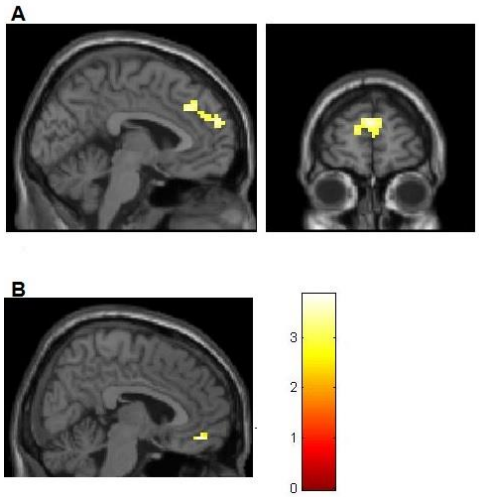

Abbildung 4: Stärkeres BOLD-Signal in der Gruppe der Proselfs während sozialer Entscheidungen im Vergleich zu eigennützigen Entscheidungen (NCS > PSI) im dmPFC (A) und mOFC (B). A: MNI: X = -3, Z = 22. B: X = 6; Schwellenwert bei $t>2.73$, entspricht $p<0.005$.

Der vmPFC ist assoziiert mit Kosten-Nutzen-Abwägung (Basten, Biele, Heekeren, & Fiebach, 2010; de Quervain, Fischbacher, Treyer, Schellhammer, & Fehr, 2004) sowie mit subjektivem Belohnungswert (Bartra et al., 2013; Rangel & Hare, 2010). Die Entscheidungen des vorliegenden Paradigmas erforderten eine Integration der subjektiven Bewertung für den eigenen Gewinns, sowie für den Gewinn einer anderen Person und induzierten somit eine prosoziale Wertberechnung. Die vmPFC-Aktivierung war in der NCS-Bedingung besonders stark. Zudem war die Aktivierung des vmPFC in der NCS-Bedingung in Proselfs stärker ausgeprägt als in Prosocials. Dies legt nahe, dass Proselfs bei sozialen Entscheidungen verstärkt kognitive Ressourcen benötigen, um den egosistischen Default zu überwinden und einen zusätzlichen sozialen Aspekt in dem Valuationsprozess zu integrieren. Prosocials hingegen taten dies automatisierter, mit weniger kognitiven Ressourcen.

In ähnlichem Ausmaß ist die dmPFC Aktivierung als ein Korrelat kognitiver Kontrolle zu interpretieren, die benötigt wird, um prosoziale Entscheidungen zu treffen. Der dmPFC wird mit kontrollierten Formen sozialer Kognition in Verbindung gebracht (Satpute & Lieberman, 2006). Der dmPFC ist darüber hinaus bei unterschiedlichsten Aufgaben mit der Aufrechterhaltung nicht-automatisierter kognitiver Prozesse in Verbindung gebracht worden, z.B. bei verstärkter kognitiver Anforderungen während Hypothesentesten (Ferstl & von Cramon, 2002) und während Kohärenzverarbeitung von Sprache (Berthoz, Armony, Blair, & Dolan, 2002). In diesem Sinne kann die Aktivierung des dmPFC während prosozialer Entscheidungen in unserem Paradigma als ein Indikator reflektiver kognitiver Prozesse gesehen werden, welche in egoistischen Probanden verstärkt ausgeprägt waren. DmPFC ist außerdem mit *Theory of Mind* (ToM) assoziiert (Saxe, 2006). Eine stärkere Aktivierung des dmPFC in egoistischen Probanden könnte zudem eine stärkere Anforderung an Prozesse wiederspiegeln, die mit ToM assoziiert sind.

Wie andere Studien zuvor liefern diese Ergebnisse Hinweise auf die Relevanz kontrollierter, kognitiver Prozesse während prosozialer Entscheidungen (Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006). Zudem zeigten sich neuronale Korrelate intrinsischer Belohnung während Entscheidungen, die eine andere Person begünstigen ohne Kosten für den Akteur zu verursachen.

Darüber hinaus fanden wir interindividuelle Unterschiede neuronaler Korrelate prosozialer Präferenzen: Egoistische Probanden überwinden ihren egoistischen Default, um prosozial zu handeln und dies ging mit Aktivierungen in Arealen einher, die zum einen mit kontrollierter sozialer Kognition (dmPFC) assoziiert sind, zum anderen mit der Integration von Entscheidungs- und Belohnungswerten (vmPFC). Die Ergebnisse erlauben eine detailliertere Sicht auf die Frage, ob prosoziales Verhalten intuitiv ist oder deliberativ mit der Unterdrückung egoistischer Impulse einhergeht, indem interindividuelle Unterschiede Berücksichtigung finden. Durch diese Studie wird deutlich, dass es nicht eine Frage von

entweder oder ist, sondern es abhängig von dem Persönlichkeitsmerkmal der prosozialen Wertorientierung ist: Proselfs handeln reflektiv und deliberativ und benötigen kognitive Ressourcen, um prosozial zu handeln, während prosoziale Personen eher intuitiv prosozial handeln. Wir finden sowohl neuronale, als auch behaviorale Indikatoren, die dies stützen.

2.2 Die Charity-Studie

Das Spenden für wohltätige Zwecke ist eine besondere Form prosozialen Verhaltens, da es Kosten für den Spender verursacht, ohne eine Gegenleistung zu erhalten und exklusiv unbekannten Dritten zugutekommt. In der Ökonomie werden unterschiedliche, dem Spendenverhalten zugrundliegende Motive diskutiert (Konow, 2010). Diese lassen sich grob in zwei Klassen unterteilen: handlungsbezogene und ergebnisorientierte Motive (siehe dazu auch Kapitel 1.2). Ergebnisorientierte Spendenmotivation setzt voraus, dass die Person eine tatsächliche Präferenz für die Höhe der Spende hat, unabhängig davon, wie die Spende zustande kommt. In diesem Fall beinhaltet das Spendengeld an sich einen Belohnungswert, unabhängig von der Höhe der *eigenen* Spende. In diesem Zusammenhang sprechen wir von ergebnisorientierten Motiven (*outcome-orientation* in der englischen Version der Veröffentlichung). Andreoni (1990) nennt diese Motivation *pure altruism*. Davon abzugrenzen sind Motive, die mit der Handlung des Spendens verknüpft sind. Hier spielt die Höhe der Spende eine untergeordnete Rolle. Es entsteht ein angenehmes Gefühl durch die Handlung an sich, ein *warm-glow of giving* (Andreoni, 1990). In diesem Zusammenhang sprechen wir von handlungsbezogenen Motiven (*action-orientation* in der englischen Version der Veröffentlichung).

Sowohl Verhaltens- (Andreoni, 1990; Konow, 2010) als auch fMRT-Studien (Harbaugh et al., 2007; Moll et al., 2006) lassen den Schluss zu, dass beide Motivklassen bei Spendenentscheidungen relevant sind. Sowohl Moll und Kollegen, als auch Harbaugh und Kollegen fanden Aktivierung im dopaminergen Belohnungssystem während Spendenentscheidungen.

Unsere Charity-Studie untersuchte die Relevanz handlungs- beziehungsweise ergebnisorientierter Motive bei Spendenentscheidungen mithilfe funktioneller Magnetresonanztomografie. Wir nutzten ein Konstrukt, das seine Wurzeln in der Lerntheorie hat, um auf innovative Art ergebnisorientierte Motive zu überprüfen: den Belohnungsvorhersagefehler (im Folgenden wird der englische Begriff *reward prediction error*, RPE (Schultz, 1998) verwendet). In Entscheidungsexperimenten ist es sehr schwer, die Effekte handlungs- und ergebnisorientierter Motive zu trennen, da in den meisten Experimenten Handlung und Ergebnis simultan auftreten. Der RPE ermöglicht eine direkte

Überprüfung ergebnisorientierter Motive und ist frei von entscheidungsassoziierten Prozessen.

Der RPE wurde in unserer Studie induziert, indem im Anschluss an die Entscheidung über die Verteilung von Geldbeträgen zwischen Proband und einer Spendenorganisation ein Teil der Entscheidungen verworfen wurde[3]. RPEs treten in Situationen auf, in denen Belohnungen unsicher sind und nicht vorhergesagt werden können (Yacubian, 2006). Dies war in unserem Paradigma zu dem Zeitpunkt der Fall, wenn die Entscheidung eines Probanden entweder bestätigt oder verworfen wurde (siehe Abbildung 1). Es entstanden zwei RPEs: einer für den eigenen Geldbetrag, einer für den Geldbetrag der Spendenorganisation. RPEs kodieren ein neuronales Signal, welches der Differenz zwischen tatsächlicher und erwarteter Belohnung entspricht (Schultz, 1998). Der RPE ist positiv, wenn die erhaltene Belohnung größer ist als erwartet, er ist negativ, wenn diese kleiner ist als erwartet (Yacubian, 2006). Die genaue Definition des RPEs ist in der Veröffentlichung zu finden.

Für eigene Belohnungen finden sich im dopaminergen Belohnungssystem, v.a. im Nucleus Accumbens, zuverlässig neuronale Signale, die mit der Höhe des RPEs korrespondieren (Pagnoni et al., 2002). Wir erwarteten ein solches Signal in unserer Studie zu finden, welches mit der Höhe des RPE für das eigene Geld korrespondiert. Das Finden eines RPEs für die Spendenorganisation wäre ein neuronaler Indikator für das Bestehen ergebnisorientierter Motive bei Spendenentscheidungen. Konkret nehmen wir an, dass Menschen, die bereit sind auf eigene Gewinne zu verzichten, um Geld zu spenden, dem Spendengeld an sich einen Belohnungswert beimessen und folglich für Spendengelder in ähnlicher Weise ein neuronales RPE-Signal in Belohnungsarealen zeigen, wie für eigenes Geld. Neben dem NAcc, wurden weitere, mit Belohnungsverarbeitung im sozialen Kontext assoziierte, Areale untersucht: subgenuales Areal (Moll et al., 2006) und mOFC (Hare et al., 2010).

Zudem wurden zum Zeitpunkt der Entscheidung neuronale Korrelate handlungsbezogener Spendenmotive in den genannten Belohnungsarealen untersucht.

Das Paradigma entspricht dem Paradigma des modifizierten Diktatorspiels der SVO-Studie (eine kurze Beschreibung des Paradigmas findet sich in der Zusammenfassung der SVO-Studie, Kapitel 2.1; eine ausführliche Beschreibung ist der Publikation zu entnehmen). Es besteht lediglich eine Erweiterung bezüglich der RPE-Manipulation. Denn im Falle dieser Studie fand eine zweistufige RPE-Manipulation statt: Neben der RPE-Induktion direkt im Anschluss an die Entscheidung (äquivalent zum SVO-Paradigma) fand eine weitere, zeitlich

[3] Die gewählten Alternativen wurden entweder bestätigt (50%), d.h. sie kamen in den Pool aus denen am Ende des Experiments eine zufällig gezogen und ausgezahlt wurde, oder die Alternative wurde verworfen (50%), d.h. sie hatte keine Chance für die Auszahlung realisiert zu werden.

spätere RPE-Induktion, losgelöst von der Entscheidung, statt (Details siehe Veröffentlichung).

Bezüglich der Ergebnisse ist zu berichten, dass sich behavioral eine starke Varianz in der Bedingung zeigte, in der Probanden auf eigene Kosten Geld spenden konnten (Spende in der *costly-donation* Bedingung). Einige Probanden waren oft bereit, eigene monetäre Vorteile aufzugeben, um der Spendenorganisation mehr Geld zukommen zu lassen. Andere Probanden hingegen wählten in diesen Situationen immer die Alternative, die mit einem höheren eigenen Geldgewinn einhergeht. Auf der Grundlage des Verhaltens in dieser Situation unterteilten wir die Stichprobe in sogenannte *Donator* (n=16) und *Non-Donator* (n=17) (siehe Veröffentlichung für genaue statistische Begründung dieses Schrittes).

Zum Zeitpunkt der Wahl einer Alternative zeigte sich keine spezifische neuronale Aktivierung während Spendenentscheidungen in den genannten Belohnungsarealen. Die Studie lässt also keine Aussage bezüglich handlungsbezogener Spendenmotive zu.

Während der RPE-Induktion zeigte sich eine positive Korrelation der NAcc-Aktivierung und des RPE für den eigenen Geldbetrag. In der Gruppe der *Donator* zeigte sich zudem eine positive Korrelation der NAcc-Aktivierung mit der Höhe des RPE für den Spendenbetrag (ROI-Analyse: Parameter gemittelt über alle Voxel des NAcc: $t(14) = 4.644$, $p = 0.00018$). In der Gruppe der *Non-Donator* war dieser Zusammenhang nicht zu finden (ROI-Analyse: Parameter gemittelt über alle Voxel des NAcc: $t(16) = 1.195$, $p = 0.125$). Der Unterschied zwischen den beiden Gruppen war signifikant ($t(30) = 2.164$, $p = 0.02$). Die Ergebnisse sind in Abbildung 5 zusammengefasst. Zudem zeigte sich eine positive signifikante Korrelation zwischen diesem neuronalem RPE-Signal und der Rate der Spendenentscheidungen in der *costly donation*-Bedingung (Spearman's rho = 0.309, $p = 0.043$).

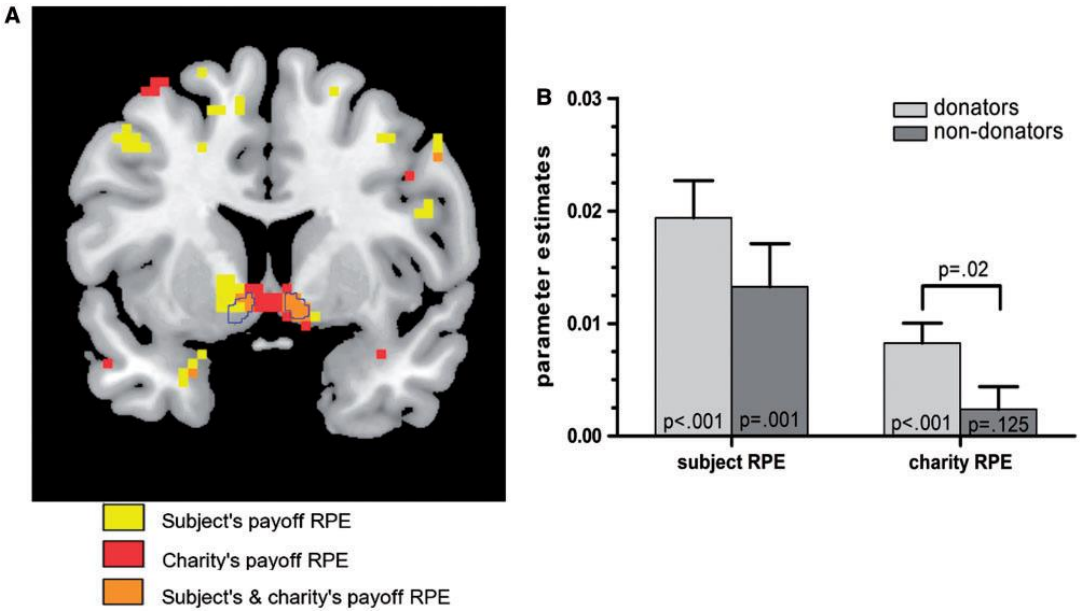

Abbildung 5: Nucleus accumbens signalisiert den Vorhersagefehler für eigenes Geld in allen Probanden („subject's payoff RPE") und spezifisch den Vorhersagefehler für Spendengelder („charity's payoff RPE") in Probanden, die bereit waren, eigene monetäre Vorteile zugunsten der wohltätigen Organisation aufzugeben (sog. *Donator*). (A) Modulation des BOLD-Signals durch den RPE für eigenes Geld in gelb für die Gesamtstichprobe (Schwellenwert bei t>5.99, entspricht p<0.0001) und durch den RPE für Spendengelder in rot für die Gruppe der *Donator* (Schwellenwert bei t>2.98, entspricht p<0.005). MNI: Y=8. Der Nucleus Accumbens ist durch blaue Ränder dargestellt. Der Effekt des Charity's-RPE ist signifikant nach small volume Korrektur für den anatomisch definierten Nucleus Accumbens (pFWE-corrected < 0.05). (B) Diagramme zeigen Aktivierungsmaße gemittelt über den bilateralen Nucleus Accumbens (± *SEM*) getrennt für *Donator* und *Non-Donator*. P's für Einstichproben T-Test gegen Null für jeden Regressor, sowie eines Zwei-Stichproben T-Tests sind zu sehen.

Die RPE-assoziierte Belohnungsaktivierung konnten wir folglich für eigenes Geld sowie für Spendengelder im NAcc finden, einem Areal, das in zahlreichen Studien mit Belohnungsverarbeitung allgemein und im Besonderen mit dem RPE in Verbindung gebracht wurde (Knutson & Cooper, 2005; Pagnoni et al., 2002). Interessant ist die genaue Lokalisation dieser Aktivierungsmuster: Das Signal des Spenden-RPE lag etwas medial außerhalb der anatomisch definierten NAcc-Maske, Richtung subgenuales Areal. Diese Region steht in Zusammenhang mit sozialen Aspekten belohnender Ereignisse (Hsu, Anen, & Quartz, 2008; Moll et al., 2006), sowie sozialer Bindung (Krueger et al., 2007) und wurde auch in Zusammenhang mit Spendenentscheidungen berichtet (Moll et al., 2006). Die Aktivierungs-Peaks des RPE des eigenen Geldbetrags lagen etwas lateraler und innerhalb der NAcc-Maske.

Eine Konjunktions-Analyse[4] bestätigte überlappende Aktivierung für RPE des eigenen Geldes und des Spendengeldes im NAcc. Es fand sich in keinem anderen Areal des Gehirns eine solche überlappende Aktivierung. Neben der überlappenden NAcc-Aktivierung für individuellen- und Spenden-RPE, fanden wir Aktivierung im mOFC, die mit dem individuellen RPE assoziiert war. Hier fanden wir keine Assoziierungen mit dem Spenden-RPE.

Die Ergebnisse unterstützen die Hypothese, dass sozialer und nicht-sozialer Kognitionen ähnliche neuronale Mechanismen zugrunde liegen (Adolphs, 2003; Ruff & Fehr, 2014): Wir konnten ähnliche neuronale Belohnungssignale für eigenes sowie für Spendengelder finden. Konkret fanden wir, dass NAcc-Aktivierung als Reaktion auf Spendengelder in der Gruppe der *Donator* in ähnlicher Weise durch den RPE parametrisch modulierbar[5] ist, wie dies für individuelle Belohnungen der Fall ist. Zudem fanden wir einen interindividuellen Unterschied: Nicht alle Probanden zeigten diese RPE-assoziierte Aktivierung für Spendengelder, sondern nur diejenigen, die auch behavioral ihre Spendenbereitschaft demonstrierten. Dies unterstützt vorherige Ergebnisses, die nahelegen, dass NAcc-Aktivierung den subjektiven Wert einer Belohnung wiederspiegelt (Tobler, Fletcher, Bullmore, & Schultz, 2007). Die Ergebnisse können als ein neurophysiologisches Korrelat ergebnisorientierter Motive bei Spendenentscheidungen interpretiert werden: Menschen, die bereit sind zu spenden, messen Spendengeldern einen Belohnungswert in ähnlicher Wiese bei, wie sie dies für eigenes Geld tun.

Neben diesen Erkenntnissen bezüglich neuronaler Grundlagen prosozialen Verhaltens, bietet die Studie auch einen methodischen Beitrag. Unsere Manipulation des RPE erlaubt es, selektiv ergebnisorientierte Motive, losgelöst von Entscheidungen, zu untersuchen. Dies ist methodisch interessant, da üblicherweise Entscheidung und Ergebnis per se miteinander verknüpft sind. Die Einführung eines RPE könnte in Entscheidungsexperimenten jeglicher Art interessant sein, um Ergebnis-Werte (outcome values) experimentell zu untersuchen.

2.3 Die Effort-Studie

Diese Studie thematisierte streng genommen nicht soziale Präferenzen in der eigentlichen Bedeutung (definiert als Entscheidungen, die die Auswirkung auf andere Personen berücksichtigen), weil in dieser Studie keine Entscheidungen über die Verteilung von Geldbeträgen getroffen wurden. Stattdessen thematisierte die Studie den Einfluss von Kontextfaktoren, die für die Erforschung sozialer Präferenzen relevant sind. Konkret

[4] Eine Konjunktions-Analyse ermöglicht es überlappende Aktivierung von 2 Kontrasten darzustellen. Es werden Voxel signifikant, in denen beide Kontraste das definierte Signifikanzniveau erreichen (logisches „und“).

[5] Parametrische Modulation beschreibt die Assoziation von BOLD-Signal und einem parametrischen Modulator (hier die Höhe des RPEs). Die Assoziation ist analog einer Korrelation zu interpretieren.

untersuchte die Studie den Einfluss von Anstrengung, die für den Erhalt eines Geldbetrags zu erbringen ist, auf die daran anschließende Belohnungsverarbeitung im Kontext von Spendenentscheidungen. Verhaltensökonomische Studien legen nahe, dass der subjektive Wert eines Geldbetrags steigt, wenn dieser durch eine anstrengende Aufgabe verdient wurde (Muehlbacher & Kirchler, 2009): Probanden gaben weniger ihres durch hohe Anstrengung verdienten Geldes ab, im Vergleich zu Geld, welches durch eine weniger anstrengende Aufgabe erzielt wurde.

Im Tierreich erhalten Individuen selten Belohnungen ohne vorherige Anstrengung. Ökologische Theorien gehen davon aus, dass die aufgebrachte Anstrengung bei dem anschließenden Prozess der Belohnungsverarbeitung berücksichtigt wird, um zukünftige Entscheidungen anzupassen (Kolling, Behrens, Mars, & Rushworth, 2012; Stephens & Anderson, 2001). Es steht also außer Frage, dass es ökologisch sinnvoll ist, die aufgebrachte Anstrengung in die Bewertung einer Belohnung miteinzubeziehen, um beispielsweise zu ermitteln, ob sich die Anstrengung gelohnt hat.

Sowohl Tier- (Rudebeck, Walton, Smyth, Bannerman, & Rushworth, 2006) als auch Humanstudien (Prévost, Pessiglione, Météreau, Cléry-Melin, & Dreher, 2010) legen nahe, dass Anstrengung als ein Kostenfaktor berücksichtigt wird und folglich den Belohnungswert einer Option reduziert („*effort discounting*“). Dies ist vor dem Treffen der Entscheidung der Fall, während verschiedene Entscheidungsalternativen gegeneinander abgewogen werden. Davon abzugrenzen ist der Effekt von vorheriger Anstrengung auf die darauffolgende Belohnungsverarbeitung. Es ist anzunehmen, dass Anstrengung in diesem Fall den hedonistischen Wert einer Belohnung steigert (Tierstudien: Clement, Feltus, Kaiser, & Zentall, 2000; Johnson & Gallagher, 2011; Humanstudien: Alessandri, Darcheville, Delevoye-Turrell, & Zentall, 2008; Zink, Pagnoni, Martin-Skurski, Chappelow, & Berns, 2004).

Die vorliegende Studie untersuchte den Effekt, den Anstrengung auf die Verarbeitung einer Belohnung hat, wenn die Anstrengung vor dem Erhalt der Belohnung zu erbringen ist. Wir nehmen an, dass Anstrengung weder einen konstanten reduzierenden, noch einen konstanten erhöhenden Effekt auf die Bewertung der Belohnung hat. Stattdessen gehen wir davon aus, dass Anstrengung differentielle Effekte, je nach Belohnungshöhe, hat: Relativ geringe Belohnungen werden devaluiert, während der subjektive Wert großer Belohnungen steigt. In anderen Worten: Die Sensitivität bezüglich der Belohnungshöhe steigt im Anschluss an eine Anstrengung. Dieser Mechanismus erscheint adaptiv, um zu bewerten, ob sich die Anstrengung gelohnt hat.

Wir untersuchten diesen Effekt sowohl auf den Erhalt einer Belohnung als auch auf den Verlust (wenn ein Teil der Belohnung weggenommen wird). Um dies zu realisieren, führten wir den Verlust von Geld in Form einer Spende ein: Die Probanden erhielten eine Belohnung

in Form eines Geldbetrags, wenn sie Rechenaufgaben richtig lösten. Wir manipulierten das Ausmaß der Anstrengung, indem die Schwierigkeit der Rechenaufgaben variierte (leichte versus schwierige Aufgaben), so dass sich drei experimentelle Bedingungen ergaben: Schwierige Rechenaufgaben entsprechen der *high-effort* Bedingung, leichte Rechenaufgaben entsprechen der *low-effort* Bedingung und vorgelöste Rechenaufgaben entsprechen der *no-effort* Bedingung als Kontrollbedingung. Nach dem Erhalt der Belohnung induzierten wir eine sogenannte „erzwungene Spende“. Dabei wurde die Belohnung zwischen Proband und einer Spendenorganisation aufgeteilt (Abbildung 6). Dies geschah zufällig und ohne Entscheidung des Probanden. Somit erlaubte dieses Paradigma, den Einfluss von Anstrengung sowohl auf Belohnungsverarbeitung (Erhalt der Belohnung nach gelöster Rechenaufgabe) als auch auf Verlustverarbeitung (erzwungene Spende) zu untersuchen. Konkret nahmen wir an, dass in der *high-effort* Bedingung, die neuronale Aktivität in belohnungs- bzw. verlustverarbeitenden Hirnregionen stärker mit Belohnungs- bzw. Verlusthöhe assoziieret ist, als in der *low-* und *no-effort* Bedingung.

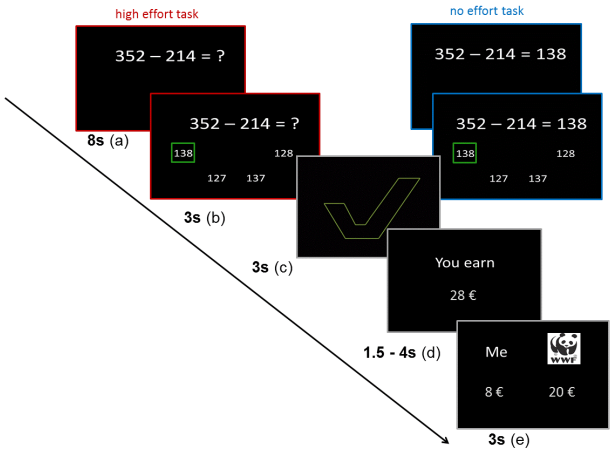

Abbildung 6: Das Paradigma der Effort-Studie. Die Probanden werden mit Rechenaufgaben unterschiedlicher Schwierigkeit konfrontiert bzw. mit bereits gelösten Rechenaufgaben (Kontroll-Bedingung) (a). Die Probanden wählen eine Lösung aus 4 Lösungsmöglichkeiten (b). Feedback bezüglich der Lösung (c). Nur im Falle der korrekten Lösung der Rechenaufgabe erhielten die Probanden einen Geldbetrag (d). Der Geldbetrag wurde durch die Experimentsoftware zufällig zwischen Proband und Spendenorganisation aufgeteilt (e).

Zum Zeitpunkt der Belohnungspräsentation nach gelöster Rechenaufgabe zeigte sich in der Bedingung der schwierigen Aufgaben (*high effort*) eine signifikante positive Modulation des BOLD-Signals mit der Höhe der Belohnung in subgenualen Region (X = 6, Y = 23, Z = -23; t = 4, small volume corrected) und im NAcc (X = 9, Y = 11, Z = -5; t = 3,5, small volume corrected). In den anderen Bedingungen zeigte sich kein solcher Effekt. Zudem zeigte sich

ein Unterschied in diesem Zusammenhang zwischen der *high effort* und *no effort* Bedingung (*high effort* > *no effort*) in beiden Arealen: subgenualen Region (X = 3, Y = 20, Z = -23; t = 4.53, small volume corrected), NAcc (X = 6, Y = 11, Z = -5; t = 2.8, small volume corrected). Der Kontrast *high effort* > *low effort* wurde nicht signifikant. Es zeigte sich kein solcher Effekt im mOFC.

Abbildung 7 verdeutlicht überlappende Aktivierungen für die differentiellen Kontraste high effort > no effort und *high effort* > *low effort* in subgenualen Region.

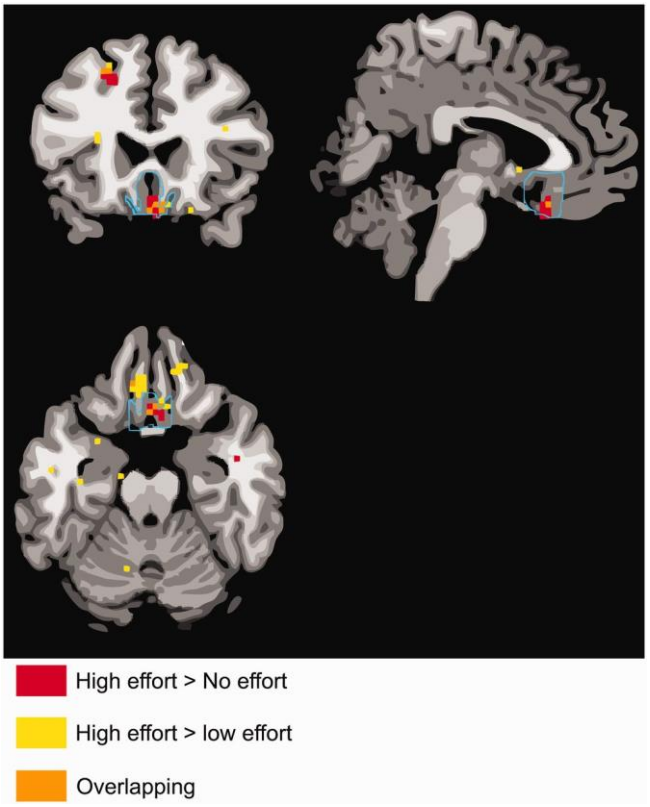

Abbildung 7: Stärkere Modulation der Aktivität in subgenualen Region (in blau) durch die Höhe des Gewinns nach starker Anstrengung (*high effort*). Es werden überlappende Aktivierungen (orange) für die differentiellen Kontraste *high effort* > *no effort* (rot, Schwellenwert bei $t > 2.67$, entspricht $p < 0.005$, unkorrigiert) und high effort > low effort (gelb, Schwellenwert bei $t > 2.39$, entspricht $p < 0.01$, unkorrigiert) gezeigt. MNI-Koordinaten: X = 3, Y = 23, Z = -21.

Wir fanden ein vergleichbares Ergebnismuster für das Verlustereignis während der erzwungenen Spende in der anterioren Insel: Eine signifikante positive Modulation des BOLD-Signals durch den relativen Verlust nur in der *high effort* Bedingung (anteriore Insel: X = -51, Y = 11, Z = 4; t = 4.53, whole brain cluster level; 106 voxel). Auch hier zeigte sich ein differentieller Effekt der Bedingungen (*high* > *low effort*: X = -51, Y = 11, Z = 4; t = 5.07,

whole brain cluster level; 189 voxel; und *high* > *no effort*: X = 39, Y = 8, Z = -8; t = 4.47, small volume corrected), siehe Abbildung 8.

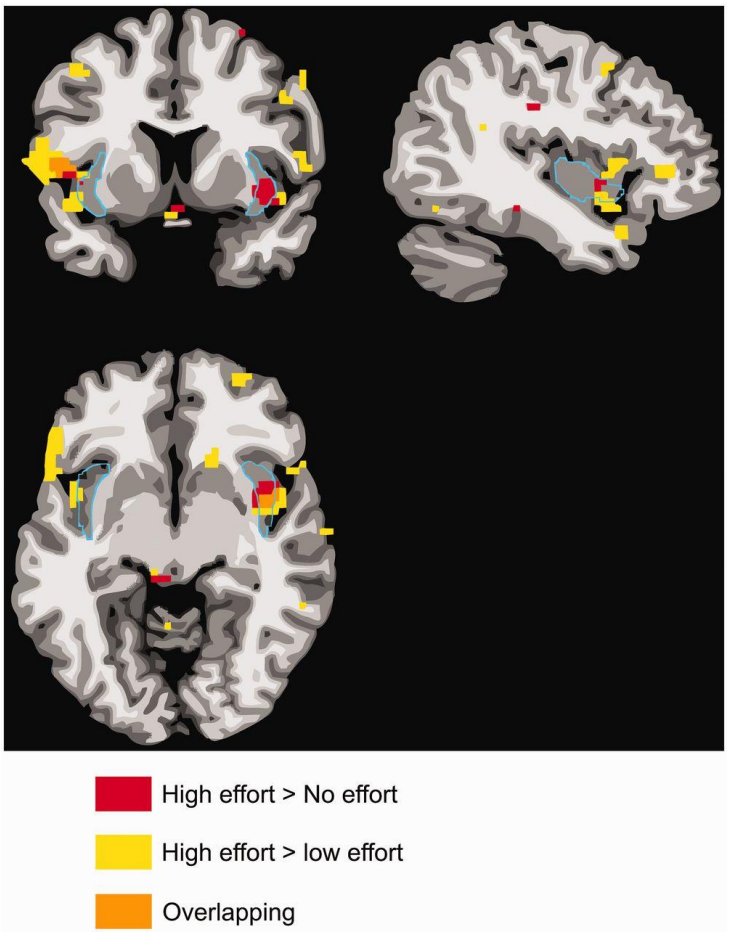

Abbildung 8: Stärkere Modulation der Aktivität in der Insel (in blau) durch die Höhe des Verlusts nach starker Anstrengung (*high effort*). Es werden überlappende Aktivierungen (orange) für die differentiellen Kontraste *high effort* > *no effort* (rot, Schwellenwert bei $t > 2.67$, entspricht $p < 0.005$, unkorrigiert) und *high effort* > *low effort* (gelb, Schwellenwert bei $t > 2.49$, entspricht $p < 0.01$, unkorrigiert) gezeigt. MNI-Koordinaten: X = 42, Y = 9, Z = -8.

Diese Ergebnisse demonstrieren eine stärkere Assoziation des BOLD-Signals mit der Höhe der Belohnung in belohnungsverarbeitenden Arealen (subgenuale Region und NAcc) nach starker Anstrengung. Zudem zeigte sich zum Zeitpunkt des Verlusts des verdienten Geldes durch die erzwungene Spende eine Assoziation der neuronalen Aktivität mit der Höhe des Verlustes in der anterioren Insel. Dies war zum einen spezifisch nach schwierigen Rechenaufgaben *(high effort)*, zum anderen stärker nach schwierigen Rechenaufgaben im Vergleich zu einfachen (*low effort)* bzw. gelösten Rechenaufgaben (*no effort)*.

Somit zeigte sich ein Effekt der Belohnungshöhe ausschließlich für Belohnungen, die durch starke Anstrengung „verdient“ wurden. Komplementär trat ein Verlustsignal für diese „verdienten“ Gelder auf. Diese Ergebnisse verdeutlichen, dass die Umstände des Erhalts einer Belohnung, sowohl die neuronale Belohnungs- als auch Verlustverarbeitung im menschlichen Gehirn beeinflussen.

Diese Ergebnisse lassen nicht den Schluss zu, dass die neuronale Aktivierung nach starker Anstrengung allgemein höher ist. Es handelt sich hierbei um die Assoziation der Höhe der Belohnung bzw. des Verlustes mit dem BOLD-Signal (starke Assoziation nach starker Anstrengung; geringere Assoziation nach geringer bzw. keiner Anstrengung). Dies könnte einen neuronalen Indikator einer stärkeren Sensitivität der Probanden gegenüber Belohnung bzw. Verlust nach Anstrengung darstellen.

Aktivität in der anterioren Insel ist mit negativen Emotionen assoziiert (Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006; Seymour et al., 2005). Im Kontext von Entscheidungsfindung zeigte sich ein negativer Zusammenhang der Inselaktivität mit dem erwarteten Belohnungswert (Rolls, Grabenhorst, & Parris, 2008) und ein positiver Zusammenhang mit der Erwartung eines aversiven Stimulus (Seymour et al., 2004). Wir interpretieren unser insuläres Signal während der erzwungenen Spende als einen neuronalen Indikator der Aversivität, die mit dem Verlust nach starker Anstrengung einhergeht. Die erzwungene Spende umfasste neben dem Verlust des eigenen Geldes auch einen Gewinn für die Spendenorganisation. Probanden, die Spendengeldern einen hohen Wert beimessen, sollten folglich auch dem Spendenereignis einen positiven Wert beimessen (in ähnlicher Weise, wie wir dies in der Charity-Studie zeigen konnten). Wir haben in dem fMRT-Experiment jedoch nicht die sozialen Präferenzen der Probanden erfasst und können diesen Faktor nicht kontrollieren. Wir fanden keinerlei Belohnungsaktivierung zum Zeitpunkt der erzwungenen Spende. Im Gesamten scheint also während der erzwungenen Spende der Verlust des eigenen Geldes gegenüber dem Gewinn der Spendenorganisation zu dominieren. Dies stimmt überein mit Ergebnissen von Harbaugh und Kollegen (2007), die eine höhere NAcc-Aktivierung während freiwilliger Spenden fanden im Vergleich zu erzwungenen Spenden.

Die Ergebnisse verdeutlichen, dass Belohnungen besonders relevant werden, wenn zuvor eine Anstrengung zum Erhalt der Belohnung getätigt wurde. Dies ist in Übereinstimmung mit ökologischen Theorien, in dem Sinne, dass ein Organismus besonders aufmerksam den Effekt einer Handlung verfolgt, wenn dieser im Zusammenhang damit viel Energie investiert hat (Kolling et al., 2012). Die Tatsache, dass neuronale Korrelate dieses Effekts experimentell durch eine simple Manipulation induziert werden können, legt nahe, dass dies ein fundamentaler Mechanismus mit starker biologischer Fundierung ist. Diese Erkenntnisse sollten bei der Gestaltung von Verhaltensexperimenten, in denen Entscheidungen bezüglich Geldes getroffen werden, Berücksichtigung finden.

3. Diskussion

Ist prosoziales Verhalten belohnend? Die in dieser kummulativen Dissertation vorgelegten Studien möchten sich mittels der Interpretation neuronaler Korrelate dieser Frage nähern. Dazu wurde ein verhaltensökonomisches Paradigma entwickelt, welches den Entscheidungsprozess in Phasen unterteilt und es dadurch ermöglicht, die Phasen des *value-based-decision makings* im Kontext sozialer Entscheidungen näher zu beleuchten (vor, während und im Anschluss an eine prosozialen Entscheidung). Wir fanden Aktivierung in belohnungsassoziierten Arealen während prosozialer Entscheidungen für eine andere Person im vmPFC (SVO-Studie), sowie im Anschluss an eine Entscheidung im NAcc (Charity-Studie). Zu diesem Zeitpunkt variierte die Aktivität im NAcc mit der Höhe der Spendengelder in der Charity-Studie in ähnlicher Wiese, wie dies für eigene Gelder der Fall ist. Diese Ergebnisse sprechen dafür, dass bei Entscheidungen, von denen andere profitieren, neuronale Prozesse relevant sind, die typischerweise bei Belohnungen im individuellen Kontext zu beobachten sind.

Neben diesen belohnungsassoziierten Aktivierungen fanden sich zum Zeitpunkt der Entscheidung in der SVO-Studie auch Aktivierungen in Arealen, die mit kognitiver Kontrolle und Deliberation assoziiert sind (dmPFC). Die prosozialen Entscheidungen erfordern ein Abwägen und Integrieren von Kosten und Nutzen, welche über eine reine Belohnungsverarbeitung hinausgehen. Somit zeigten sich in der SVO-Studie, neben den belohnungsassoziierten Aktivierungen, neuronale Indikatoren weiterer kognitiver Prozesse, im Sinne der Kontrolle über primär eigensinnige Motive während prosozialer Entscheidungen.

Die Effort-Studie knüpfte an diese Ergebnisse an und erweiterte diese um den Aspekt der Leistungserbringung bei prosozialen Entscheidungen, d.h. konkret inwieweit die Tatsache, ob jemand für einen Geldbetrag eine Leistung erbracht hat, oder nicht (*Windfall-Money*), die neuronalen Reaktionen auf diese Geldbeträge verändert. Die Ergebnisse verdeutlichen, dass das menschliche Gehirn Kontextfaktoren, bzw. die Umstände des Erhalts einer Belohnung, kodiert: Es zeigte sich eine stärkere Assoziation der Aktivierung in Belohnungsarealen (NAcc) mit der Höhe des „verdienten“ Geldes, wenn das Geld durch das Lösen einer anstrengenden Aufgabe geschah. In ähnlicher Weise zeigte sich eine Assoziation mit der Höhe des Verlustes dieses Geldes in der anterioren Insel. Diese Ergebnisse sprechen dafür, dass das Gehirn tatsächlich den *subjektiven* Belohnungswert kodiert und neben dem absoluten Wert einer Belohnung Kontextfaktoren berücksichtigt. Vorherige Studien konnten zeigen, dass die Hirnaktivierung von anderen Kontextfaktoren beeinflusst wird, wie z.B. die Belohnung einer anderen Person (Fliessbach et al., 2007) oder das Ausgangsgehalt einer anderen Person (Tricomi et al., 2010).

Die Ergebnisse im Kontext des *value-based decision making* sprechen für *common currency* und liefern Erkenntnisse bezüglich individueller Differenzen

Die SVO-Studie präsentierte Ergebnisse zum Zeitpunkt der Entscheidung. Übertragen auf das Modell des *value-based decision making*, entspricht dieser Zeitpunkt dem Entscheidungswert (*decision value*; Fehr & Krajbich, 2014). Zu diesem Zeitpunkt werden Kosten und Nutzen einer Alternative bewertet und integriert, um die Option zu wählen, die den höchsten Entscheidungswert hat (Rangel et al., 2008) bzw. am besten zur Erreichung eigener Ziele und zur Umsetzung eigener moralischer Werte dient (Gęsiarz & Crockett, 2015). Die Entscheidungen in der SVO-Studie erforderten die Integration des Wertes, den man dem eigenen Geldbetrag beimisst, und des Wertes, den man dem Geldbetrag der anderen Person beimisst, in einen subjektiven Wert. Wir fanden Aktivierung in Arealen, die typischerweise im individuellen Kontext an diesem Prozess beteiligt sind (vmPFC, mOFC; Bartra et al., 2013; Basten et al., 2010; de Quervain et al., 2004; Rangel & Hare, 2010), sowie Aktivierung im dmPFC, als Korrelat nicht-automatisierten Verhaltens von prosozialen Entscheidungen (Satpute & Lieberman, 2006). Neben den klassischen, belohnungsassoziierten *Value*-Arealen, waren mit dem dmPFC folglich auch Areale beteiligt, die mit kognitiver Kontrolle, Arbeitsgedächtnis und kontrollierter, sozialer Kognition in Verbdingung gebracht wurden (Elliott & Dolan, 1998; Ferstl & von Cramon, 2002; Satpute & Lieberman, 2006). Dies verdeutlicht den komplexen Charakter dieser Entscheidung: Kosten und Nutzen, in diesem Falle, prosozialer Entscheidungen werden gegeneinander abgewogen. Dieses neuronale Aktivierungsmuster kann als Korrelat einer prosozialen Wertberechnung gesehen werden.

Zudem liefert die Studie Erkenntnisse bezüglich interindividueller Differenzen zwischen Probanden mit unterschiedlichem Grad der sozialen Wertorientierung (*social value orientation*; Van Lange, 1999) und bezüglich der Frage, ob prosoziales Verhalten automatisiert ist oder mit der Kontrolle egoistischer Impulse einhergeht. Wir fanden stärkere Aktivierungen während prosozialer Entscheidungen in der Gruppe der egoistischen Probanden im Vergleich zu den prosozialen Probanden. Die Bedingung, in der diese Aktivierung beobachtet wurde, hatte ein besonderes Merkmal: Es war die einzige Bedingung, in der eigengewinn-maximierende Entscheidungen nicht gemacht werden konnten. In dieser Bedingung wurden egoistische Probanden dazu gebracht ihren egoistischen *Default*, der darin besteht lediglich die Konsequenzen für ihre eigene Auszahlung zu berücksichtigen, zu überwinden und eine prosoziale Entscheidung zu treffen. Die stärkere Aktivierung des vmPFC in der Gruppe der Proselfs während dieser Entscheidungen wird als Korrelat eines aufwändigeren Valuations-Prozesses interpretiert, im Vergleich zu intuitiveren prosozialen Entscheidungen in der Gruppe der Prosocials. In ähnlicher Weise ist die stärkere dmPFC-Aktivierung der egoistischen Probanden als ein Korrelat des reflexiven Prozesses zu sehen,

der kognitive Ressourcen (und folglich weniger Automatisierung) bedarf. DmPFC ist ebenfalls mit *Theory of Mind* (ToM) sowie der Verarbeitung sozial relevanter Stimuli assoziiert (Saxe, 2006). In diesem Sinne kann der Gruppenunterschied in der dmPFC-Aktivierung als ein Korrelat höherer Anforderungen sozialer Kognition in der Gruppe der Proselfs (ToM, mentalizing) interpretiert werden.

Die Gruppenunterschiede in den Reaktionszeiten ergänzen die neuronalen Ergebnisse und sprechen dafür, dass egoistische Probanden ihren egoistischen Default überwinden, um prosozial zu handeln: Egoistische Probanden brauchten am längsten für konfliktfreie prosoziale Entscheidungen, da sie in dieser Bedingung nicht einfach die Alternative mit der höheren eigenen Auszahlung wählen konnten. Prosocials hingegen entschieden in der konflikthaften sozialen Bedingung am langsamsten, wenn sie die egoistische Alternative wählten. Neuronale und behaviorale Indikatoren stimmen überein.

Die Ergebnisse erlauben durch die Berücksichtigung interindividueller Unterschiede eine detailliertere Sicht auf die Frage ob prosoziales Verhalten intuitiv ist oder deliberativ mit der Unterdrückung egoistischer Impulse einhergeht. Dabei handelt es sich nicht um eine Frage von Entweder-oder, sondern es ist abhängig von Persönlichkeitsmerkmalen: Proselfs handeln reflektiv und deliberativ und benötigen kognitive Ressourcen, um prosozial zu handeln, während prosoziale Personen eher intuitiv und automatisiert prosozial handeln.

Diese Sichtweise findet sich ebenso in der Unterscheidung von Verhaltenssystemen in *habitual-* und *goal-directed behavior* (Gęsiarz & Crockett, 2015): Verhalten des *goal-directed-behavior* -Systems beansprucht kognitive Ressourcen und kann durch Übung in habituelles Verhalten übergehen, welches weniger kognitive Ressourcen bedarf und automatisiert abläuft. Prosoziales Verhalten der Prosocials ist tendenziell eher dem *habitual-System* zuzuordnen, während prosoziales Verhalten der Proselfs eher die Kriterien des *goal-directed behavior* aufweist. Hervorzuheben ist der Fakt, dass durch Übung und durch Verstärkung prosoziales Verhalten gelernt und habituell werden kann (Gęsiarz & Crockett, 2015). Hier ist eine Parallele zu der sozialen Lerntheorie Banduras (1977) zu sehen: Prosoziales Verhalten wird zunächst gezeigt, weil es durch externe Belohnungen verstärkt wird, bis es schließlich, über einen weiteren Schritt der sozialen Verstärkung (z.B. Lob), zu internalen Belohnungen führt und an dieser Stelle als habituell bezeichnet werden kann. Auf dieser letzten Entwicklungsstufe geht prosoziales Verhalten mit internalen Selbstbelohnungen einher und ist vermutlich automatisiert und teilweise habituell, wie das Verhalten der Prosocials. Hier ist eine wechselseitige Beeinflussung von stabilen, internalen, moralischen Werten (wie der prosozialen Wertorientierung), und dem Einfluss der Umwelt anzunehmen.

In der Charity-Studie fanden sich ebenfalls Hinweise für interindividuelle Unterschiede prosozialen Verhaltens: Personen, die bereit waren, eigene Gewinne aufzugeben, um der wohltätigen Organisation im Experiment Geld zu spenden, wurden als sog. *Donator*

klassifiziert. Diese Probanden zeigten einen Belohnungsvorhersagefehler (RPE) bezüglich der Spendengelder, was wir als einen neuronalen Indikator des ultimativen Interesses an der Spendenhöhe deuten. Denn die Aktivität im NAcc variierte mit der Höhe der Spendengelder in ähnlicher Wiese, wie dies für eigene Gelder der Fall ist, für die gewiss ein ultimatives Interesse besteht. In der Gruppe der *Non-Donator* konnten wir ein solches Signal nicht detektieren. Während in der SVO-Studie die Probanden anhand eines Persönlichkeitsmerkmals vorab in Gruppen eingeteilt wurden, wurde in der Charity-Studie das im Experiment gezeigte Verhalten für die Gruppenbildung genutzt. Das Einteilen der Probanden nach Persönlichkeitsmerkmalen im Vorlauf des fMRT-Experiments bietet einige Vorteile: Es konnten gezielter a priori Hypothesen zu diesen Gruppen untersucht werden. Zudem waren wir nicht darauf angewiesen, das Entscheidungsverhalten, das den Gruppenunterschied bedingt, selbst im fMRT-Experiment zu evozieren, was immer die Gefahr des Zirkelschlusses birgt.

Unsere Ergebnisse deuten darauf hin, dass soziale und nicht-soziale Kognitionen gemeinsame neuronale Verarbeitungsmechanismen nutzen (entsprechend der *common currency* Hypothese, Ruff & Fehr, 2014). Während prosozialer Entscheidungen für eine andere Person in der SVO-Studie waren Areale aktiv, die in zahlreichen Studien während Entscheidungen im individuellen Kontext (vmPFC, mOFC) gefunden wurden (z.B. Bartra et al., 2013; Rangel & Hare, 2010). Das Finden überlappender Aktivierung für den Belohnungsvorhersagefehler für eigenes Geld sowie für Spendengeld *innerhalb einer Studie (*Charity-Studie*)*, ist ein weiterer, und vor allem direkterer, Beleg für die *common currency*-Hypothese.
Wir fanden jedoch auch Hinweise für spezifische soziale Aktivierungsmuster, die die *social valuation specific Hypothese* stützen (Ruff & Fehr, 2014): In der Charity-Studie ragte die Aktivierung für das Spendengeld bis in die septale Region. Dabei handelt es sich um eine Region, die mit sozialer Bindung assoziiert ist (Krueger et al., 2007). Neben der vorwiegenden Überlappung der beiden Aktivierungen war mit der spezifischen Aktivierung der septalen Region für den Spenden-RPE auch eine Spezialisierung zu finden. Es ist ein hervorzuhebendes Merkmal der Charity-Studie, dass neuronale Reaktionen auf egoistische und prosoziale Geldbeträge innerhalb einer Studie beobachtet wurden und somit ein direkter Test der überlappenden Aktivierungen im Sinne der *common currency* Hypothese möglich war.
Es ist wichtig bezüglich der *common currency* Hypothese anzumerken, dass auf Ebene einzelner Neurone ein unterschiedliches Bild herrschen kann. Es ist möglich, dass innerhalb einer Region, einzelne Neurone spezifisch auf soziale Stimuli reagieren und somit auf Ebene einzelner Neurone die *social-valuations specific* Hypohtese zutrifft (Ruff & Fehr, 2014).

Aufgrund der geringen räumlichen Auflösung des MRTs lassen die Studien keine Aussage darüber zu.

Es ist durchaus möglich, dass beide Hypothesen (*common currency* und *social specific*) zutreffen, dass also soziale und nicht-soziale Kognition sowohl die gleichen Areale aktivieren, als auch spezifische Areale der Verarbeitung sozialer Kognition dienen. Sehr plausibel scheint zu sein, dass die Areale des Valuation-Netzwerkes (in erster Linie der vmPFC) domänenübergreifend aktiviert werden, jedoch Input aus Arealen erhalten, die spezifisch mit sozialer Kognition assoziiert sind; z.B. TPJ. So zeigten Hare und Kollegen (2010) in einem Spenden-Paradigma, dass der vmPFC Signale des sozialen Kognitions-Netzwerks integriert. Die Studien der vorliegenden Dissertation können dazu keine Aussagen treffen, da keine Konnektivitätsanalysen durchgeführt wurden.

Neuronale Indikatoren ergebnis- und handlungsorientierter Motive prosozialen Verhaltens

Die Studien liefern Hinweise für die Relevanz, sowohl ergebnis- als auch handlungsorientierter Motive während prosozialer Entscheidungen (siehe Kapitel 1.2).

Das in der Charity- und SVO-Studie benutzte experimentelle Paradigma bietet die Möglichkeit der Überprüfung ergebnisorientierter Motive mittels des Belohnungsvorhersagefehlers. Somit ermöglichte das Paradigma eine weitere Komponente des Entscheidungsprozesses des *value-based decison makings* genauer zu untersuchen: Im Anschluss an die Entscheidung wurde ein Teil der Entscheidungen „verworfen“, ein anderer Teil wurde „bestätigt“. Diese Manipulation erzeugte einen Vorhersagefehler (RPE) für den eigenen Geldgewinn und den Geldgewinn für die Spendenorganisation. Diese Manipulation isolierte neuronale Reaktionen auf die Geldbeträge und war frei von entscheidungsassoziierten neuronalen Aktivierungen. Dadurch war die Überprüfung ergebnisorientierter Motive möglich.

Die Charity-Studie zeigt die Relevanz ergebnis-orientierter Motive bei Spendenentscheidungen: Die Aktivität im NAcc variierte mit der Höhe der Spendengelder in ähnlicher Wiese, wie dies für eigene Gelder der Fall ist. Hier zeigte sich jedoch, dass dies nur für Personen der Fall war, die bereit waren, eigene Gewinne aufzugeben, um der wohltätigen Organisation Geld zu spenden („Donator“). Die Ergebnisse legen nahe, dass zumindest für einen Teil der Probanden, ein ultimatives Interesse an der Höhe der Spende besteht, in ähnlicher Weise, wie dies für eigenes Geld der Fall ist.

Entscheidungsassoziierte Aktivität im vmPFC *während prosozialer Entscheidungen*, wie in der SVO-Studie beschrieben, legt die Effektivität der Handlungsorientierung nahe. Diese Aktivierung könnte ein Korrelat des *warm glow of giving* sein (Konow, 2010). Menschen ziehen aus dem Akt des Gebens einen Belohnungswert. Die SVO- und die Charity-Studie

verwendeten das gleiche Paradigma mit unterschiedlichen Rezipienten: zum einen eine Spendenorganisation, zum anderen eine real existierende Person. In der Charity-Studie fanden sich mit dem Belohnungsvorhersagefehler ergebnisorientierte Korrelate, in der SVO-Studie fanden sich mit den belohnungsassoziierten Aktivierungen zum Zeitpunkt der Entscheidung handlungsorientierte Korrelate. Dies lässt die spekulative Vermutung zu: Bei Spendenentscheidungen ist die Höhe der Spende relevant, während bei Entscheidungen für eine konkrete Person der Akt des Gebens belohnend ist. Der Umstand, dass der Spenden-RPE vor allem in der Gruppe der Spender gefunden wurde, ist in diesem Zusammenhang plausibel: Diese Probanden drücken sowohl in ihrem Verhalten als auch neuronal ein Interesse an der Höhe der Spende aus und zeigen behaviorale und neuronale Indikatoren ergebnisorientierter Motive. Im Gegensatz dazu könnte die konkrete, real existierende Person in der SVO-Studie die Möglichkeit einer Interaktion mit dieser Person in den Vordergrund treten lassen, denn die Probanden waren Studenten und hatten eine gewisse Ähnlichkeit miteinander. Darin könnte die Akzentuierung auf die Handlung der prosozialen Entscheidung begründet sein.

Diese Aktivierung in belohnungsassoziierten Hirnarealen während prosozialer Entscheidungen und in Assoziation mit der Spendenhöhe in Anschluss an eine Entscheidung verdeutlicht, dass Prosozialität mit einem neuronalen Belohnungskorrelat einhergeht und legt den verlockenden Schluss nahe, dass Prosozialität belohnend ist (siehe jedoch reverse Inferenzen).

Die Studien erlauben jedoch keine Aussage zu der in der Psychologie geführten Debatte, ob die Motivation, wirklich altruistisch ist, in dem Sinne, dass das ultimative Ziel der Handlung das Wohlergehen der anderen Person ist (Batson & Shaw, 1991). Egoistische Motive für prosoziale Entscheidungen sind nicht gänzlich auszuschließen. In dem Fall ist die prosoziale Handlung nur ein Instrument, um das eigene Wohlergehen zu steigern, z.B. Anerkennung durch andere zu erhalten. Das verwendete Paradigma wurde gestaltet, um einige dieser egoistischen Motive prosozialen Verhaltens zu kontrollieren (Anonymität, keine Interaktionen im Sinne der Abhängigkeit der Auszahlung). Dennoch sind nicht alle egoistischen Motive kontrollierbar. Prosoziales Verhalten kann zu internalen Belohnungen beim Erreichen von Zielen und beim Umsetzen eigener moralischer Werte führen, z.B. dem Wert eine gute Person zu sein und zu der Bestätigung des Selbstbildes einer „guten Person“ führen (Gęsiarz & Crockett, 2015). Diese, im Grunde ebenfalls egoistische Motivation, ist nicht zu kontrollieren.

Methodische Aspekte

Das Paradigma der Entscheidungsstudien (SVO- und Charity-Studie) bietet eine gewisse Eleganz in der Untersuchung interindividueller Differenzen prosozialer Entscheidungen. Um neuronale Korrelate interindividueller Differenzen zu untersuchen, ist es vorteilhaft, neuronale Aktivierungsunterschiede bei identischem Verhalten zu finden. Es liegt jedoch in der Natur der Variabilität prosozialen Verhaltens, dass einige Personen oft, andere hingegen kaum prosoziales Verhalten zeigen (Bogaert et al., 2008; Declerck et al., 2013). Dies stellt die Erforschung prosozialen Verhaltens vor eine Herausforderung, da egoistische Personen nur selten prosozial handeln, v.a. in ökonomischen Paradigmen. Die *non-costly social* Bedingung ist so gestaltet, dass eine hohe Rate prosozialer Entscheidungen auftritt, und es somit ermöglicht Unterschiede neuronaler Aktivierungsmuster bei identischem Verhalten zu untersuchen.

Zudem teilte das Paradigma, angelehnt an Theorien des *value-based decision makings* (Rangel & Hare, 2010), den Prozess der Entscheidung in einzelne Phasen auf: Neben der eigentlichen Entscheidung wurden vor und nach der Entscheidung Valuation-Signale getestet. In Entscheidungsexperimenten ist es sehr schwer, die Effekte handlungs- und ergebnisorientierter Motive zu trennen, da in den meisten Experimenten Handlung und Ergebnis simultan auftreten. So geht der Erhalt einer Belohnung mit der Entscheidung, die zu der Belohnung führt, zeitlichen eng zusammen. Dies macht es schwer, die neuronalen Reaktionen bezüglich der Entscheidung und bezüglich der Belohnung zu trennen. Die Charity-Studie bedient sich eines Konstruktes aus der Lerntheorie, um neuronale Reaktionen auf einen Belohnungsstimulus frei von entscheidungsassoziierten Kognitionen zu erfassen: den RPE. Durch eine einfache experimentelle Manipulation (eine Entscheidung zu bestätigen bzw. zu verwerfen) wird ein RPE induziert.

Die Effort-Studie nutzte ebenfalls eine methodische Finesse, Reaktionen auf Belohnungswerte, frei von Entscheidungen, zu untersuchen. Nachdem die Probanden einfache Rechenaufgaben, schwierige Rechenaufgaben oder Rechenaufgaben mit vorgegebener Lösung (Kontrollaufgaben) bearbeitet hatten, erhielten sie Geld. Anschließend wurde ein Teil des so erzielten Geldgewinns an eine Spendenorganisation überführt. Auf diese Geldaufteilung hatten die Probanden keinen Einfluss: Die Aufteilung geschah automatisch durch die Experiment-Software. Die Ergebnisse zeigten eine signifikante Modulation der anschließenden Belohnungsverarbeitung durch unsere experimentelle Manipulation, sowohl zum Zeitpunkt des Erhalts des Geldes im NAcc als auch zum Zeitpunkt des Verlustes in der anterioren Insel. Dies könnte einen neuronalen Indikator einer stärkeren Sensitivität der Probanden gegenüber Belohnung bzw. Verlust nach Anstrengung darstellen.

Die Tatsache, dass sich ein derartiger „Verdiensteffekt“ durch eine einfache experimentelle Manipulation erzeugen lässt, hat unserer Einschätzung nach erhebliche inhaltliche und

methodische Konsequenzen für das Gebiet der Verhaltensökonomie und die kognitiven Neurowissenschaften, die sich mit der Verarbeitung von Belohnungen beschäftigen. In verhaltensökonomischen Experimenten ist es üblich, den Probanden im Rahmen des Experiments Geld zu geben, mit dem sie im Experiment agieren (z.B. in sozialen Experimenten aufteilen; aber auch in Börsen-Experimenten investieren). Im Anschluss an das Experiment wird den Probanden das erwirtschaftete Geld üblicherweise ausgezahlt. Der Vorteil dieses Vorgehens ist die Realitätsnähe. Zu bedenken ist jedoch, dass dieses Geld in gewisser Weise „vom Himmel fällt" (in diesem Zusammenhang wird auch von *Windfall Money* gesprochen). Unsere Ergebnisse liefern Hinweise, dass es einen Unterschied macht, ob das Geld „vom Himmel fällt" oder verdient wird. Dies zeigte sich nicht nur neuronal, sondern in der Tendenz auch im Verhalten (kleinere Spenden nach schwieriger Rechenaufgabe; siehe Publikation). Der Mensch ist besonders sensibel bezüglich der Verarbeitung von Geldbeträgen, wenn er zuvor etwas dafür geleistet hat. Dies ist auch aus ökologischer Perspektive sinnvoll: wurde viel Energie für eine Handlung aufgewandt, ist es besonders relevant, ob sich diese Anstrengung auch gelohnt hat, beispielsweise im Tierreich bei Nahrungssuche (Clement et al., 2000; Johnson & Gallagher, 2011).

Die Ergebnisse stimmen ebenfalls mit der Annahme überein, dass Attribution von Belohnungen auf internale oder externale Ursachen die anschließende Bewertung beeinflusst. Dies wird von der Attributionstheorie thematisiert (Weiner, 2000) und auf Verhaltensebene bestätigt (Wittig, Marks, & Jones, 1981). Die Ergebnisse sind auch mit dem psychologischen Konstrukt *Deservingness* in Verbindung zu bringen. *Deservingness* ist definiert als Balance zwischen Belohnung und Handlung, die zur Belohnung führt (Feather & McKee, 2009; Feather, McKee, & Bekker, 2011). Es ist anzunehmen, dass die erhöhte neuronale Sensitivität bezüglich Belohnungs- und Verlusthöhe nach Anstrengung neuronal zu diesem *Deservingness*-Mechanismus beiträgt.

Die Effort-Studie verdeutlicht, dass eine einfache Manipulation die neuronale Verarbeitung von Geldbeträgen verändern kann und liefert Hinweise, dass Kontextfaktoren, in diesem Fall der Umstand des Erhalts einer Belohnung, mitkodiert werden. Dieser Effekt lässt sich mit Konstrukten unterschiedlicher Wissenschaften in Verbindung bringen, u.a. Attributionstheorien (Weiner, 2000) sowie dem Konzept *Deservingness* (Feather et al., 2011) in der Psychologie; *Windfall Money* (Muehlbacher & Kirchler, 2009) in der Ökonomie; Nahrungssuche-Theorien (Kolling et al., 2012) in der Biologie.

Limitationen

Die Natur von fMRT-Ergebnissen ist korrelativ: Es wird ein zeitlicher Zusammenhang zwischen Blutfluss im Gehirn und Aufgabenmerkmalen gestellt. Es ist folglich nicht eindeutig

zu schlussfolgern, dass die gemessene Hirnaktivierung eine kausale Rolle in dem beobachteten, prosozialen Verhalten spielt. Läsionsstudien (Krajbich, Adolphs, Tranel, Denburg, & Camerer, 2009; Stone et al., 2010) sowie Studien, die neuronale Aktivität experimentell beeinflussen können, wie z.B. transkranielle magnetische Stimulation (Knoch et al., 2006; Strang et al., 2012), legen jedoch eine kausale Rolle des dorsolateralen PFC und Arealen des Valuations-Netzwerks bei prosozialen Entscheidungen nahe. Diese Aussage ist für die hier präsentierten Ergebnisse jedoch nicht zulässig.

FMRT-Ergebnisse verleiten zu reversen Inferenzen (Englisch, *reverse inference*; Poldrack, 2006): Ist ein Areal aktiv, schreibt man diesem die gleichen mentalen Prozesse zu, die in anderen Studien in Zusammenhang mit diesem Areal beschrieben werden. Auf der Suche nach erklärbaren, konkreten kognitiven Prozessen ist dieser Fehler bei den sonst sehr abstrakten Aktivierungen einzelner (und mitunter zahlreicher) Areale im fMRT schnell gemacht. Diese Schlüsse sind jedoch nicht zulässig: Ein Areal kann an vielen verschiedenen kognitiven Prozessen beteiligt sein. Das Zuschreiben einer bestimmten Kognition *aufgrund* der Aktivierung engt den Blickwinkel ein und ist inkorrekt, da in den allermeisten Fällen Schlüsse von der Aktivierung auf den kognitiven Prozesses nicht möglich sind. Auch die vorliegenden Arbeiten bedienen sich reverser Interferenzen. Die Aktivierungen in NAcc, vmPFC und mOFC werden genutzt, um auf zugrundeliegende psychologische Konstrukte, in diesem Fall auf Präferenzen und Motive, zu schließen. Es gibt Hinweise darauf, dass Belohnungsaktivierungen im NAcc als Surrogatmarker individueller Präferenzen gesehen werden kann (Knutson, Delgado, & Phillips, 2008). Derartige Schlüsse müssen jedoch mit Vorsicht gezogen werden, da Hirnaktivierungen im MRT immer mehrdeutig sind.

Literaturverzeichnis

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nat.Rev.Neurosci.*, *4*(1471–003X (Print)), 165–178. http://doi.org/10.1038/nrn1056

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annu Rev Psychol*, *60*, 693–716. http://doi.org/10.1146/annurev.psych.60.110707.163514

Alessandri, J., Darcheville, J.-C., Delevoye-Turrell, Y., & Zentall, T. R. (2008). Preference for rewards that follow greater effort and greater delay. *Learning & Behavior*, *36*(4), 352–8. http://doi.org/10.3758/LB.36.4.352

Andreoni, J. (1990). IMPURE ALTRUISM AND DONATIONS TO PUBLIC GOODS : A THEORY OF WARM-GLOW GIVING. *The Economic Journal*, *100*(401), 464–477.

Bandura, Albert (1977). Social Learning Theory. Oxford, England: Prentice-Hall.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427. http://doi.org/10.1016/j.neuroimage.2013.02.063

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21767–21772. http://doi.org/10.1073/pnas.0908104107

Batson, C. D., Dyck, J. L., Brandt, J. R., Batson, J. G., Powell, A. L., McMaster, M. R., & Griffitt, C. (1988). Five studies testing two new egoistic alternatives to the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, *55*(1), 52–77. http://doi.org/10.1037/0022-3514.55.1.52

Batson, C. D. & Powell, A. A. (2003). Altruism and Prosocial Behavior. In I. B. Weiner, H. A. Tennen, J. M. Suls (Eds.), Handbook of Psychology, Volume 5, Personality and Social Psychology , pp. 463–484. New Jersey: John Wiley & Sons, Inc.

Batson, D., & Shaw, L. (1991). Evidence for Altruism: Toward a Pluralism of Prosocial Motives. *Psychological Inquiry*, *2*(2), 107–122.

Bereczkei, T., Birkas, B., & Kerekes, Z. (2010). Altruism towards strangers in need: costly signaling in an industrial society. *Evolution and Human Behavior*, *31*(2), 95–103. http://doi.org/10.1016/j.evolhumbehav.2009.07.004

Berthoz, S., Armony, J. L., Blair, R. J. R., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain*, *125*(8), 1696–1708. http://doi.org/10.1093/brain/awf190

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: a review and conceptual model. *The British Journal of Social Psychology*, *47*, 453–480. http://doi.org/10.1348/014466607X244970

Boone, C., Declerck, C., & Kiyonari, T. (2010). Inducing Cooperative Behavior among Proselfs versus Prosocials: The Moderating Role of Incentives and Trust. *Journal of Conflict Resolution*, *54*, 799–824. http://doi.org/10.1177/0022002710372329

Bray, S., & O'Doherty, J. (2007). Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *Journal of Neurophysiology*, *97*(4), 3036–3045. http://doi.org/10.1152/jn.01211.2006

Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., & Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron*, *30*(2), 619–39.

Camerer, C.F. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. Princeton NJ: Princeton University Press.

Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The*

Quarterly Journal of Economics, *117*(3), 817–869. http://doi.org/10.1162/003355302760193904

Cialdini, R. B., Schaller, M., Houlihan, D., Arps, K., Fultz, J., & Beaman, A. L. (1987). Empathy-based helping: Is it selflessly or selfishly motivated? *Journal of Personality and Social Psychology*, *52*(4), 749–758. http://doi.org/10.1037/0022-3514.52.4.749

Clement, T. S., Feltus, J. R., Kaiser, D. H., & Zentall, T. R. (2000). "Work ethic" in pigeons: reward value is directly related to the effort or time required to obtain the reward. *Psychonomic Bulletin & Review*, *7*(1), 100–6.

Cornelissen, G., Dewitte, S., & Warlop, L. (2011). Are social value orientations expressed automatically? Decision making in the dictator game. *Personality and Social Psychology Bulletin*, *37*(8), 1080–1090. http://doi.org/10.1177/0146167211405996

Darley, J. M., & Latane, B. (1968). BYSTANDER INTERVENTION IN EMERGENCIES: DIFFUSION OF RESPONSIBILITY. *Journal of Personality and Social Psychology*, *8*(4), 377–383.

de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., & Fehr, E. (2004). The Neural Basis of Altruistic Punishment. *Science*, (305), 1254–1258.

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: An fMRI investigation. *NeuroImage*, *23*(2), 744–751. http://doi.org/10.1016/j.neuroimage.2004.05.025

Declerck, C. H., & Bogaert, S. (2008). Social value orientation: Related to empathy and the ability to read the mind in the eyes. *The Journal of Social Psychology*, *148*(6), 711–726. http://doi.org/10.3200/SOCP.148.6.711-726

Declerck, C. H., Boone, C., & Emonds, G. (2013). When do people cooperate? The neuroeconomics of prosocial decision making. *Brain and Cognition*, *81*(1), 95–117. http://doi.org/10.1016/j.bandc.2012.09.009

Dovidio, J. E., Allen, J., & Schroeder, D. A. (1990). The specificity of empathy-induced helping: Evidence for altruism. *Journal of Personality and Social Psychology*, *59*, 249–260.

Elliott, R., & Dolan, R. J. (1998). Activation of different anterior cingulate foci in association with hypothesis testing and response selection. *NeuroImage*, *8*(8), 17–29. http://doi.org/10.1006/nimg.1998.0344

Emonds, G., Declerck, C. H., Boone, C., Vandervliet, E. J. M., & Parizel, P. M. (2011). Comparing the neural basis of decision making in social dilemmas of people with different social value orientations, a fMRI study. *Journal of Neuroscience, Psychology, and Economics*, *4*(1), 11–24. http://doi.org/10.1037/a0020151

Erk, S., Spitzer, M., Wunderlich, A. P., Galley, L., & Walter, H. (2002). Cultural objects modulate reward circuitry. *Neuroreport*, *13*(18), 2499–2503. http://doi.org/10.1097/00001756-200212200-00024

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, *54*(2), 293–315. http://doi.org/10.1016/j.geb.2005.03.001

Feather, N. T., & McKee, I. R. (2009). Differentiating emotions in relation to deserved or undeserved outcomes: A retrospective study of real-life events. *Cognition & Emotion*, *23*(5), 955–977. http://doi.org/10.1080/02699930802243378

Feather, N. T., McKee, I. R., & Bekker, N. (2011). Deservingness and emotions: Testing a structural model that relates discrete emotions to the perceived deservingness of positive or negative outcomes. *Motivation and Emotion*, *35*(1), 1–13. http://doi.org/10.1007/s11031-011-9202-4

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*(7208), 1079–1083. http://doi.org/10.1038/nature07155

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences*, *11*(10), 419–427. http://doi.org/10.1016/j.tics.2007.09.002

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–791. http://doi.org/10.1038/nature02043

Fehr, E., & Krajbich, I. (2014). Social Preferences and the Brain. In P. W. Glimcher & E. Fehrt (Eds.), *Neuroeconomics – Decision Making and the Brain.* (2nd ed., pp. 193–218). Oxford: Elsevier Inc.

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness , Competition , and Cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.

Ferstl, E. C., & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: coherence or theory of mind? *NeuroImage*, *17*(3), 1599–1612. http://doi.org/10.1006/nimg.2002.1247

Fliessbach, K. (2011). Soziale Präferenzen. In M. Reimann & B. Weber (Eds.), *Neuroökonomie - Grundlagen, Methoden, Anwendungen* (pp. 139–159). Wiesbaden: Springer.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., & Falk, A. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science (New York, N.Y.)*, *318*(5854), 1305–8. http://doi.org/10.1126/science.1145876

Forsythe, R., Horowitz,J., Savin, N.E. & Sefton,M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior,* 6, 347-369.

Gęsiarz, F., & Crockett, M. J. (2015). Goal-directed, habitual and Pavlovian prosocial behavior. *Frontiers in Behavioral Neuroscience*, *9*(May), 1–18. http://doi.org/10.3389/fnbeh.2015.00135

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal O F Theoretical Biology.*, *7*, 1–16. http://doi.org/10.1016/0022-5193(64)90038-4

Harbaugh, W. T. (1998). What do donations buy?: A model of philanthropy based on prestige and warm glow. *Journal of Public Economics*, *67*(2), 269–284. http://doi.org/10.1016/S0047-2727(97)00062-5

Harbaugh, W. T., Mayr, U., & Burghart, D. R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science (New York, N.Y.)*, *316*(5831), 1622–1625. http://doi.org/10.1126/science.1140738

Hare, T. A., Camerer, C. F., Knoepfle, D. T., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(2), 583–90. http://doi.org/10.1523/JNEUROSCI.4089-09.2010

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors. *Journal of Neuroscience*, *28*(22), 5623–5630. http://doi.org/10.1523/JNEUROSCI.1309-08.2008

Haruno, M., & Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*(2), 160–1. http://doi.org/10.1038/nn.2468

Haruno, M., Kimura, M. & Frith, C.D. (2014). Activity in the nucleus accumbens and amygdala underlies individual differences in prosocial and individualistic economic choices. *Journal of Cognitive Neuroscience*, 26(8):1861-70. doi: 10.1162/jocn_a_00589.

Hernandez Lallement, J.*, Kuss, K.*, Trautner, P., Weber, B., Falk, A., Fliessbach, K. (2014).

Effort increases sensitivity to reward and loss magnitude in the human brain. *Social Cognitive and Affective Neuroscience*, 9(3), 342-349. doi: 10.1093/scan/nss147. Epub 2012 Nov 30.

Hoffman, E., McCabe, K. & Smith, V.L. (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review*, 86, 653–60.

Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: distributive justice and neural encoding of equity and efficiency. *Science*, *320*(5879), 1092–1095. http://doi.org/10.1126/science.1153651

Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of Social and Monetary Rewards in the Human Striatum. *Neuron*, *58*(2), 284–294. http://doi.org/10.1016/j.neuron.2008.03.020

Johnson, A. W., & Gallagher, M. (2011). Greater effort boosts the affective taste properties of food. *Proceedings. Biological Sciences / The Royal Society*, *278*(1711), 1450–6. http://doi.org/10.1098/rspb.2010.1581

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

Kable, J. W., & Glimcher, P. W. (2009). The Neurobiology of Decision: Consensus and Controversy. *Neuron*, *63*(6), 733–745. http://doi.org/10.1016/j.neuron.2009.09.003

Kahnt, T., Heinzle, J., Park, S. Q., & Haynes, J.-D. (2010). The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(13), 6010–5. http://doi.org/10.1073/pnas.0912838107

Kampe, K. K. W., Frith, C. D., Dolan, R. J., & Frith, U. (2001). Psychology: Reward value of attractiveness and gaze. *Nature*, *413*, 589.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*(5800), 829–832. http://doi.org/10.1126/science.1129156

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *21*(16), RC159.

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology*, *18*(4), 411–7.

Knutson, B., Delgado, M. R., & Phillips, P. E. M. (2008). Representation of subjective value in the striatum. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics – Decision Making and the Brain.* (pp. 389–406). Elsevier Inc. http://doi.org/10.1016/B978-0-12-374176-9.00025-7

Knutson, B., Rick, S., Wimmer, G. E., Prelec, D., & Loewenstein, G. (2007). Neural Predictors of Purchases. *Neuron*, *53*(1), 147–156. http://doi.org/10.1016/j.neuron.2006.11.010

Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural Mechanisms of Foraging. *Science*, *336*(6077), 95–98. http://doi.org/10.1126/science.1216930

Konow, J. (2010). Mixed feelings: Theories of and evidence on giving. *Journal of Public Economics*, *94*(3–4), 279–297. http://doi.org/10.1016/j.jpubeco.2009.11.008

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., & Camerer, C. F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(7), 2188–92. http://doi.org/10.1523/JNEUROSCI.5086-08.2009

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., … Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences of*

the United States of America, *104*(50), 20084–9. http://doi.org/10.1073/pnas.0710103104

Kuss, K., Falk, A., Trautner, P., Elger, C. E., Weber, B., & Fliessbach, K. (2013). A reward prediction error for charitable donations reveals outcome orientation of donators. *Social Cognitive and Affective Neuroscience*, 8 (2), 216–223. doi: 10.1093/scan/nsr088

Kuss, K., Falk, A., Trautner, P., Montag, C., Weber, B., & Fliessbach, K. (2015). Neuronal correlates of social decision making are influenced by social value orientation - an fMRI study. *Frontiers in Behavioral Neuroscience*, *9*(February), 1–8. doi: 10.3389/fnbeh.2015.00040

Latane, B., & Darley, J. M. (1968). Group Inhibition of Bystander Intervention in Emergencies. *Journal of Personality and Social Psychology*, *10*(3), 215–221.

Latane, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin*, *89*(2), 308–324. http://doi.org/10.1037/0033-2909.89.2.308

Levy, I., Lazzaro, S. C., Rutledge, R. B., & Glimcher, P. W. (2011). Choice from non-choice: Predicting consumer preferences from BOLD signals obtained during passive viewing. *Journal of Neuroscience*, *31*(1), 118–125. http://doi.org/10.1523/JNEUROSCI.3214-10.2011.Choice

Maner, J. K., Luce, C. L., Neuberg, S. L., Cialdini, R. B., Brown, S., & Sagarin, B. J. (2002). The Effects of Perspective Taking on Motivations for Helping: Still No Evidence for Altruism. *Personality and Social Psychology Bulletin*, *28*(11), 1601–1610. http://doi.org/10.1177/014616702237586

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(42), 15623–8. http://doi.org/10.1073/pnas.0604475103

Muehlbacher, S., & Kirchler, E. (2009). Origin of Endowments in Public Good Games : The Impact of Effort on Contributions. *Journal of Neuroscience, Psychology, and Economics*, *2*(1), 59–67. http://doi.org/10.1037/a0015458

O'Doherty, J. P., Deichmann, R., Critchley, H. D., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron*, *33*(5), 815–826. http://doi.org/10.1016/S0896-6273(02)00603-7

Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience*, *5*, 97–98.

Paulus, M. P., & Frank, L. R. (2003). Ventromedial prefrontal cortex activation is critical for preference judgments. *Neuroreport*, *14*(10), 1311–5. http://doi.org/10.1097/01.wnr.0000078543.07662.02

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–5. http://doi.org/10.1038/nature05051

Plassmann, H., O'Doherty, J., & Rangel, A. (2007). Orbitofrontal Cortex Encodes Willingness to Pay in Everyday Economic Transactions. *Journal of Neuroscience*, *27*(37), 9984–9988. http://doi.org/10.1523/JNEUROSCI.2131-07.2007

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. http://doi.org/10.1016/j.tics.2005.12.004

Prévost, C., Pessiglione, M., Météreau, E., Cléry-Melin, M.-L., & Dreher, J.-C. (2010). Separate valuation subsystems for delay and effort decision costs. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(42), 14080–90. http://doi.org/10.1523/JNEUROSCI.2752-10.2010

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, *17*(8), 413–425. http://doi.org/10.1016/j.tics.2013.06.003

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. http://doi.org/10.1038/nrn2357

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology*, *20*(2), 262–70. http://doi.org/10.1016/j.conb.2010.03.001

Rawls, J. (1971). A theory of justice. Cambridge: Harvard University Press.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., & Kilts, C. D. (2002). A Neural Basis for Social Cooperation, *35*, 395–405.

Rolls, E. T., Grabenhorst, F., & Parris, B. a. (2008). Warm pleasant feelings in the brain. *NeuroImage*, *41*(4), 1504–13. http://doi.org/10.1016/j.neuroimage.2008.03.005

Rudebeck, P. H., Walton, M. E., Smyth, A. N., Bannerman, D. M., & Rushworth, M. F. S. (2006). Separate neural pathways process different decision costs. *Nature Neuroscience*, *9*(9), 1161–8. http://doi.org/10.1038/nn1756

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews. Neuroscience*, *15*(8), 549–562. http://doi.org/10.1038/nrn3776

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, *1079*(1), 86–97. http://doi.org/10.1016/j.brainres.2006.01.005

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239. http://doi.org/10.1016/j.conb.2006.03.001

Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*, *80*, 1–27.

Seymour, B., Doherty, J. P. O., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., … Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans, *429*(June), 664–667. http://doi.org/10.1038/nature02636.1.

Seymour, B., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., & Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature Neuroscience*, *8*(9), 1234–40. http://doi.org/10.1038/nn1527

Steinbeis, N., Bernhardt, B. C., & Singer, T. (2012). Impulse Control and Underlying Functions of the Left DLPFC Mediate Age-Related and Age-Independent Individual Differences in Strategic Social Behavior. *Neuron*, *73*(5), 1040–1051. http://doi.org/10.1016/j.neuron.2011.12.027

Stephens, D. W., & Anderson, D. (2001). The adaptive value of preference for immediacy : when shortsighted rules have farsighted consequences. *Behavioral Ecology*, *12*(3), 330–339.

Stone, V. E., Cosmides, L., Tooby, J., Kroll, N., Knight, R. T., & Knightlll, R. T. (2010). impairment of with in a patient exchange damage system Selective reas Dning b ilateral about limbic social. *Sciences-New York*, *99*(17), 11531–11536. http://doi.org/10.1073/pnas.

Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., & Sack, A. (2012). Be Nice if You Have to - The Neurobiological Roots of Strategic Fairness. *Social Cognitive and Affective Neuroscience*, *541*, 1–11.

Tobler, P. N., Fletcher, P. C., Bullmore, E. T., & Schultz, W. (2007). Learning-related human brain activations reflecting individual finances. *Neuron*, *54*(1), 167–75. http://doi.org/10.1016/j.neuron.2007.03.004

Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, *463*(7284), 1089–1091.

http://doi.org/10.1038/nature08785

van den Bos, W., & Guroglu, B. (2009). The Role of the Ventral Medial Prefrontal Cortex in Social Decision Making. *Journal of Neuroscience*, *29*(24), 7631–7632. http://doi.org/10.1523/JNEUROSCI.1821-09.2009

Van Lange, P. A. M. (1999). The Pursuit of Joint Outcomes and Equality in Outcomes: An Integrative Model of Social Value Orientation. *Journal of Personality and Social Psychology*, *77*(2), 337–349.

Wedekind, C., & Braithwaite, V. A. (2002). The long-term effects of human generosity in indirect reciprocity. *Current Biology*, *12*(2), 1012.

Weiner, B. (2000). Intrapersonal and Interpersonal Theories of Motivation from an Attributional Perspective. *Educational Psychology Review*, *12*(1), 1–14.

Wittig, M. a., Marks, G., & Jones, G. a. (1981). Luck versus Effort Attributions: Effect on Reward Allocations to Self and Other. *Personality and Social Psychology Bulletin*, *7*(1), 71–78. http://doi.org/10.1177/014616728171011

Yacubian, J. (2006). Dissociable Systems for Gain- and Loss-Related Value Predictions and Errors of Prediction in the Human Brain. *Journal of Neuroscience*, *26*(37), 9530–9537. http://doi.org/10.1523/JNEUROSCI.2915-06.2006

Zaki, J., & Mitchell, J. P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences*, *108*(49), 19761–19766. http://doi.org/10.1073/pnas.1112324108

Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., & Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron*, *42*(3), 509–17. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15134646

Danksagung

Ich bedanke mich bei Prof. Dr. Martin Reuter für seine Bereitschaft meine Dissertation zu betreuen und seine Unterstützung, die er mir dabei zukommen ließ, vor allem in der abschließenden Phase der Promotion.

Mein ganz besonderer Dank gilt PD Dr. Klaus Fließbach, der mich mit seiner fachlichen Kompetenz und Geduld wissenschaftliches Arbeiten in kognitiven Neurowissenschaften gelehrt hat. Danke für die Unterstützung über all die Jahre! Ich schätze es sehr mit Dir zu arbeiten!

Ich danke Prof. Dr. Michael Wagner für seine freundliche Bereitschaft im Rahmen meines Promotionsverfahrens den Prüfungsvorsitz zu übernehmen. Darüber hinaus danke ich ihm dafür, mich für die klinische Neuropsychologie zu begeistern und mich in diesem mir sehr ans Herz gewachsene Feld der Psychologie zu unterstützen.

Ich danke Prof. Dr. Ulrich Ettinger für seine freundliche Bereitschaft als weiteres prüfungsberechtigtes Mitglied an der Prüfungskommission teilzunehmen.

Ich danke den Kollegen des CENs und life & brain, ganz besonders Sabrina Strang, Niklas Häusler, Xenia Grote, Tina Strombach, Sarah Rudorf, Manuel Becker und Laura Schinabeck. Unser fachlicher und persönlicher Austausch haben die Zeit ganz besonders geprägt und schön gemacht. Ich danke Prof. Dr. Armin Falk für die Möglichkeit am CENs zu arbeiten und für seine Unterstützung von Forschungsprojekten. Darüber hinaus möchte ich mich bei Prof. Dr. Bernd Weber für die Möglichkeit bedanken, die hervorragende Infrastruktur des life & brain und des CENs nutzen zu dürfen.

Ganz besonders danke ich meiner Mutter und meiner Tochter Lina, dafür dass ihr so gut Zeit zusammen verbringen konntet und es mir ermöglicht habt an der Dissertation zu schreiben.

Meinem Ehemann Fabian danke ich für seine Geduld, sein Verständnis und dafür mich an entsprechender Stelle zu motivieren.

Ich danke meiner Familie und meinen Freunden für das wirklich Wichtige im Leben!

Anhang

Die SVO-Studie……..S 55

Kuss, K., Falk, A., Trautner, P., Montag, C., Weber, B., & Fliessbach, K. (2015). Neuronal correlates of social decision making are influenced by social value orientation - an fMRI study. *Frontiers in Behavioral Neuroscience*, *9*(February), 1–8. doi:10.3389/fnbeh.2015.00040

Die Charity-Studie……S 63

Kuss, K., Falk, A., Trautner, P., Elger, C. E., Weber, B., & Fliessbach, K. (2013). A reward prediction error for charitable donations reveals outcome orientation of donators. *Social Cognitive and Affective Neuroscience*, 8(2), 216–223. doi:10.1093/scan/nsr088.

Die Effort-Studie……S 71

Hernandez Lallement, J., Kuss, K., Trautner, P., Weber, B., Falk, A., Fliessbach, K. (2014). Effort increases sensitivity to reward and loss magnitude in the human brain. Social Cognitive and Affective Neuroscience, 9(3), 342-349. doi: 10.1093/scan/nss147. Epub 2012 Nov 30.

frontiers in
BEHAVIORAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 24 February 2015
doi: 10.3389/fnbeh.2015.00040

Neuronal correlates of social decision making are influenced by social value orientation—an fMRI study

Katarina Kuss*[1,2], *Armin Falk*[1], *Peter Trautner*[3], *Christian Montag*[4], *Bernd Weber*[1,3,5] and *Klaus Fliessbach*[1,2,6]

[1] Center for Economics and Neuroscience, University of Bonn, Bonn, Germany
[2] Department of Psychiatry, University Hospital Bonn, Bonn, Germany
[3] Life and Brain Center, Department of NeuroCognition, University Hospital Bonn, Bonn, Germany
[4] Department of Psychology, University of Ulm, Ulm, Germany
[5] Department of Epileptology, University Hospital Bonn, Bonn, Germany
[6] Clinical Research, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

Edited by:
Pablo Brañas-Garza, Middlesex University London, UK

Reviewed by:
Maria Repolles, Middlesex University London, UK
Luis Miller, University of the Basque Country, Spain

****Correspondence:***
Klaus Fliessbach, Department of Psychiatry, University Hospital Bonn, Sigmund-Freud-Strasse 25, 53105 Bonn, Germany
e-mail: Klaus.Fliessbach@ukb.uni-bonn.de

Our decisions often have consequences for other people. Hence, self-interest and other-regarding motives are traded off in many daily-life situations. Interindividually, people differ in their tendency to behave prosocial. These differences are captured by the concept of social value orientation (SVO), which assumes stable, trait-like tendencies to act selfish or prosocial. This study investigates group differences in prosocial decision making and addresses the question of whether prosocial individuals act intuitively and selfish individuals instead need to control egoistic impulses to behave prosocially. We address this question via the interpretation of neuronal and behavioral indicators. In the present fMRI-study participants were grouped into prosocial- and selfish participants. They made decisions in multiple modified Dictator-Games (DG) that addressed self- and other-regarding motives to a varying extent (self gain, non-costly social gain, mutual gain, costly social gain). Selfish participants reacted faster than prosocial participants in all conditions, except for decisions in the non-costly social condition, in which selfish participants displayed the longest decision times. In the total sample we found enhanced neural activity in the ventromedial prefrontal cortex (vmPFC) and dorsomedial prefrontal cortex (dmPFC/BA 9) during decisions that resulted in non-costly social benefits. These areas have been implicated in cognitive control processes and deliberative value integration. Decisively, these effects were stronger in the group of selfish individuals. We believe that selfish individuals require more explicit and deliberative processing during prosocial decisions. Our results are compatible with the assumption that prosocial decisions in prosocials are more intuitive, whereas they demand more active reflection in selfish individuals.

Keywords: prosocial decision making, interindividual differences, SVO, egoistic default, valuation, cognitive control

INTRODUCTION

Prosocial behavior, i.e., behavior which benefits other individuals is a relative unique human ability. There is an ongoing scientific debate to what extent prosocial behavior is based on intuition or on conscious reasoning (Fehr and Camerer, 2007). Recent studies suggest that this might strongly depend on personality traits, and that individuals with a tendency to act prosocial ("prosocials") do so intuitively, while individuals who have a tendency to act selfish, sometimes rely on their ability to deliberatively control their selfish tendencies in order to act prosocial (Bogaert et al., 2008; Declerck et al., 2013). Neurocognitive research has recently begun to identify brain processes which are related to prosocial behavior and they have found that prosocial decisions in persons with a prosocial personality trait are associated with activity in subcortical brain regions that have been implied in automated, intuitive processing (Haruno and Frith, 2010; Haruno et al., 2014). In the present fMRI study we wanted to investigate brain processes in selfish individuals which are given the opportunity to act prosocially without own costs. Under the assumption that these individuals have a weaker default tendency to act prosocial we hypothesized that they would need extra cognitive resources to do so manifesting in longer reaction times and stronger activity in brain areas that are associated with deliberative decision making such as the prefrontal cortex.

Standard models of human decisions making assume that individuals are intuitively self-interested and primarily maximize their own gain (Camerer, 2003). Therefore, individuals have to suppress egoistic impulses in order to act prosocially (Knoch et al., 2006). Compatible with this view, egoistic choices are fast (Piovesan and Wengström, 2009) and presuambly automated. On the contrary, a study by Rand and colleagues suggests that individuals are intuitively prosocial and cooperative (Rand et al.,

2012). They found that faster decisions result in higher monetary contributions to a public good than slow decisions.

Studies that take interindividual differences into account, suggest that these conflicting views might be resolved by controlling for personality traits. The concept "Social Value Orientation" (SVO) (Van Lange, 1999) considers two kinds of stable (trait-like) preferences for resource allocation: a prosocial value orientation, which refers to the preference of maximizing the sum of resources between self and other, and a proself value orientation, which refers to the preference of maximizing individual resources. Cornelissen et al. (2011) studied the effects of social value orientation on prosocial behavior in a dictator-game under cognitive load during which prosocials transferred more money to the recipient compared to selfish participants. The authors concluded that SVO determines behavior when it is based on automatization, resulting prosocials to intutiveily act in a prosocial manner and selfish individuals to intuitively act in an egoistic manner.

Additional support for this view comes from neurocognitive studies that explore differences in brain acitivity between prosocial and selfish individuals (Van den Bos et al., 2009; Haruno and Frith, 2010; Emonds et al., 2011; Haruno et al., 2014). They found evidence that behavior and cognition of prosocials are characterized by intuition and automatization. Work by Haruno and colleagues demonstrated that prosocials rely more on automatic emotional processing when responding to inequitable monetary distributions between themselves and others. They found correlates in subcortical structures such as the Amygdala and the Nucleus Accumbens (Haruno and Frith, 2010; Haruno et al., 2014). This finding was complemented by another study that provided evidence for a higher reward value of prosocial decisions in prosocial-oriented individuals: Van den Bos found higher activity in the Striatum of prosocials compared to proselfs when reciprocating in a Trust game (Van den Bos et al., 2009). Similarly Emonds et al. (2011) interpreted higher activity in the lateral orbitofrontal cortex in prosocials during decision making in a Prisoner's Dilemma as neural indicators of intuitive, internalized moral consideration. Neural indicators for deliberative strategies and control of egoistic impulses in proselfs are reported in the same study (Emonds et al., 2011). They found higher activity in the DLPFC in proselfs for decision making in the Prisoner's Dilemma. The DLPFC has been considered to be an important brain region for working memory, because it is activated when humans are under cognitive load. Therefore higher activity in this brain area might reflect high cognitive effort to fight automatic egoistic impulses.

Our fMRI study aimed at extending those findings to decision-making in a non-interactive paradigm. We are interested in the decision process that underlies the construct of social value orientation and thus implemented a binary choice task that resembles the structure of the triple dominance measure of SVO (Van Lange, 1999). Participants choose between two alternatives, each alternative consisting of a payoff for themselves and another participant (example see **Figure 1**). Our paradigm conceptually resembles a dictator-game (Engel, 2011): decisions affect payoffs of the decision maker and of another participant (the recipient), without the recipient having influence on the payoff distribution. This excludes the possible influence of strategic motives and motives to punish unfair behavior which are effective in interactive paradigms such as the Ultimatum game (Camerer, 2003).

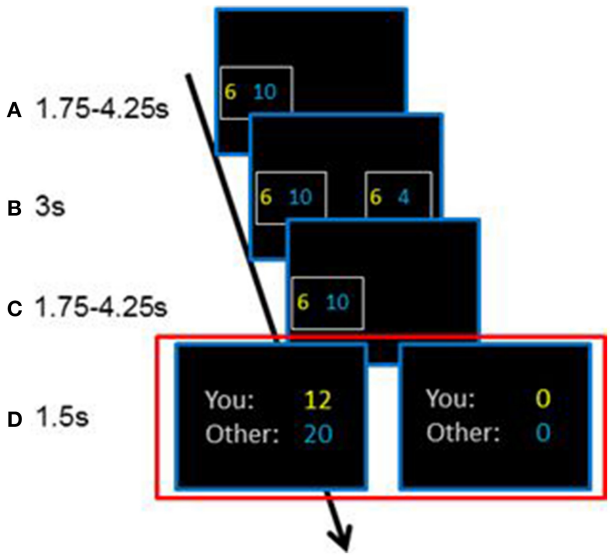

FIGURE 1 | Chronology of a trial in the paradigm. After the appearance of the first alternative **(A)**, consisting of a payoff for the decision-maker in yellow and the receiver in blue, a second alternative appeared and the subjects chose one alternative **(B)**. The chosen alternative was presented **(C)** and the chosen outcomes were either doubled or set to zero with a 50% chance **(D)**. The chronology was the same in each of the 160 trials.

Our paradigm allowed to study the valuation and integration of self-interest and prosocial motives and the relative contribution of different brain structures to prosocial decisions. We expected prosocials to act intuitively in a prosocial manner, and selfish individuals to act intuitively in an egoistic manner. The main question was how selfish individuals would react in situations in which they could act prosocially without minimizing their personal benefit. Either they don't act prosocially (at most at chance level), or they act prosocial and need additional cognitive resources to do so. To answer this question we designed situations (non-costly social condition) which focus on the payoff maximization of the other person (by holding the own payoff constant and varying the other person's payoff, e.g., choosing between 4/10 and 4/6). Since these decisions do not lead to monetary losses for the decision-maker, we expected a high rate of prosocial decisions. Decisions in the costly social condition entail a conflict between self-interest and other-regarding motives. In this situation subjects can choose to forgo a monetary advantage for themselves in order to allocate more money to the recipient (e.g., favoring 4€ for themselves and 16€ for the recipient over 6€ for both). In those situations we expected prosocials participants to behaviourally display their value for prosocial outcomes by choosing the prosocial alternative more often compared to selfish individuals.

Our study is able to shed light on the neural correlates underlying the decision process of the SVO construct and the relative contribution of brain areas involved in more automated processing (such as the Nucleus Accumbens and the Amygdala) and brain

areas involved in higher order reflective processing (such as the lateral and medial prefrontal cortex) (Satpute and Lieberman, 2006). We wanted to contribute to the question, whether individuals that value prosociality (i.e., prosocials) primarily act intuitively, and whether selfish individuals need to control egoistic impulses to behave prosocially. We expected prosocials to show stronger neural correlates of reward and valuation in subcortical structures of the brain (NAcc, Amygdala), and we expected selfish indidivuals to show stronger engagement of prefrontal areas during prosocial decisions.

METHODS

PARTICIPANTS

Prior to the fMRI experiment, each of our 40 subjects (22 female, mean age = 30.03y, SD = 8.7y) was classified as proself ($n = 20$) or prosocial ($n = 20$) based on the Social Value Orientation decomposed measure (Van Lange, 1999), which was filled out online. The SVO decomposed measure consists of 9 items. Each item contains three outcome distributions between oneself and an anonymous other, in this case another participant of the online-questionnaire (points to self/points to other). The three outcome distributions of each item correspond to a prosocial (e.g., 500/500), individualistic (e.g., 600/200) or competitive orientation (e.g., 500/0). Participants were classified as prosocial when they made at least 6 consistent prosocial choices, and classified as proself when they made at least 6 consistent individualistic choices. Competitive choices were too few to be classified. Participants were classified before the fMRI experiment to achieve an equal distribution of proself- and prosocial-oriented participants. Four subjects (all proselfs) had to be excluded from fMRI-analysis due to excessive head movement. Throughout the manuscript we use the term selfish to refer to proself individuals as defined in the SVO construct. Subjects were native German speakers, right handed and had no history of psychiatric or neurological disorders. Informed written consent was obtained from all subjects. The study was approved by the Ethics committee of the University of Bonn.

EXPERIMENTAL PROCEDURE AND PARADIGM

In each of the 160 trials of the fMRI experiment, subjects chose one of two alternatives, each consisting of a payoff for themselves and for another participant (**Figure 1**). Subjects were informed that the payoff of one randomly chosen trial would be implemented after the experiment (actual payoff). No deception was used: the selection of the implemented trial was random and the actual payoffs were transferred to the subjects (decision maker and another participant as the receiver) and subjects were guaranteed anonymity of their decisions.

Subjects were invited in groups of two and took part in the fMRI-experiment one after another. The choices the subjects made affected each other; each subject had the role of a dictator and of a receiver, thus the paradigm corresponds to a role-reversal dictator game. The randomly chosen actual payoff at the end of each experiment thus determined a payoff for the decision-maker (dictator) and a payoff for the other subject of the dyad as the receiver. The actual payoff for each participant henceforth consisted of the sum of two randomly chosen decisions: one at the end of their own experiment, the other at the end of the other participant's experiment. The subjects additionally received a show-up fee of 20 Euro. The subjects didn't know each other in advance. They briefly recognized one another before the first subject entered the scanner in order to demonstrate the receiver to actually exist. The two participants had no further contact and didn't get to know the decisions of each other. We decided to implement minimal prior contact between the subjects in order to increase ecological validity. Additional several studies demonstrated a positive influence of prior contact and of perceived similarity on cooperative behavior and its neuronal correlates (Boone et al., 2008, 2010; Mobbs et al., 2009).

The decision-process consisted of three time-points (see **Figure 1**). After the appearance of the first alternative showing a payoff for the decision-maker and the receiver, a second alternative appeared and the subjects chose one alternative. After the choice, we implemented a Reward-Prediction-Error (RPE) by either doubling the chosen outcomes or setting them to zero with a 50% chance. This allows us to test for neuronal correlates before choice (during appearance of first alternative), during choice, and after choice (RPE) (for a more detailed description of this procedure see Supplementary Material).

Subjects' and receivers' payoffs varied independently among 4, 6, 10, 16, and 20 Euros, and payoff alternatives were randomly chosen from all possible unique combinations. This led to four qualitatively distinguishable decision situations: "pure self-interest" (PSI), "non-costly social" (NCS), "efficiency" (E) and "costly social" (CS) situations (The generation of decision situations is explained in detail in the Supplementary Material). In the costly social situations, subjects could choose to forgo monetary advantages in order to allocate more money to the receiver (e.g., favoring 6€ for themselves and 16€ for the receiver over 10€ for both). In this situation, subjects had to trade material self-interest with altruistic preferences. The other three situations do not entail a conflict between different motives because one alternative is unequivocally advantageous with respect to self-interest motives (PSI), efficiency (NCS), or both motives (E). The non-conflicting nature of the NCS-condition was created by keeping the subject's payoff constant in the two alternatives and only varying the receiver's outcome (See **Table 1**: 6/10 vs. 6/4). In this situation the subjects can choose the alternative with the higher outcome for the other participant without affecting the own outcome. The PSI condition was constructed in an equivalent manner by keeping the receiver's payoff constant and varying the subject's payoff (e.g., 10/6 vs. 4/6). In the efficient condition one alternative consisted of higher payoffs for decision-maker and receiver (16/10 vs. 10/6). Subjects were presented with 40 decisions of each condition in random order, resulting in 160 decisions in total. The experimental paradigm was adopted from our previous study (Kuss et al., 2013).

In order to characterize the neuronal correlates of social decision-making, the analysis of fMRI-data concentrates on the following comparison (time-point of choice, see **Figure 1B**): the contrast between social decisions in the NCS-condition with self-interested choices in the PSI-condition (NCS > PSI). This contrast is of special interest, because conditions are formally equivalent (no conflict of motives, 1 payoff per alternative is the

Table 1 | The four decision situations and their underlying payoff-structures including percentages of trials in which subjects chose the left alternative in each condition separately for prosocials and proselfs (selfish participants).

Decision situation	Percentage of trials averaged across subjects (mean ± SD)		
	Prosocials	Proselfs	*t*-value (p)
Pure self-interest (PSI) e.g., 10/6 4/6	94.1% (± 12.92%)	95% (± 12.11%)	−0.22 (0.829)
Efficiency (E) e.g., 16/10 4/6	95.5% (± 8.96%)	95% (± 12.53%)	0.14 (0.886)
Non-costly social (NCS) e.g., 6/10 6/4	90.3% (± 16.25%)	92.1% (± 12.7%)	−0.36 (0.724)
Costly social (CS) e.g., 4/10 10/6	19.6% (± 16.32%)	6.9% (± 10.94%)	2.79 (0.008)

same) and conditions differ only in the decision's consequence with NCS-choices affecting the receiver's payoff and PSI-choices affecting the decision-maker's payoff. The events of costly social decisions (costly-social condition) were too rare to be considered in the fMRI-analysis.

TECHNICAL DETAILS

Scanning was performed on a 1.5T Avanto scanner (Siemens, Erlangen, Germany) using standard scanning parameters for the acquisition of 31 axial EPI slices with a TR of 2.5s (for details see Supplementary Material). The experiment was presented by Presentation® software version 14.9 (Neurobehavioral Systems, Albana, Canada) via video goggles (Nordic Neuro Lab, Norway) and subjects gave their answer by button presses on MRI-suited response grips (Nordic Neuro Lab).

fMRI ANALYSIS

We included three events in the first level general linear model (GLM): onset of the appearance of the first alternative including parametric modulators representing the subject's payoff and the receiver's payoff (event 1), different onset-regressors depending on the decision situations (event 2), onset of RPE-induction including two parametric modulators representing the RPE of the subject's and the receiver's payoff (event 3). Regarding event 2, the following decision-types were modeled in the GLM: In the PSI condition trials where subjects chose the self-interest alternative (PSI+), in the NCS-condition trials where subjects choose the prosocial alternative (NCS+), in the E-condition trials where subjects chose the efficient alternative (E+), in the CS-condition trials where subjects chose the prosocial alternative (CS+), and trials where subjects chose the self-interest alternative (CS−). All decisions, except prosocial decision in the costly social condition (CS+), were made with a certain frequency to be considered as a reliable regressor in the GLM (at least 20 decisions per condition).

Importantly, the parametric regressor for the others's payoff in event 1 and event 3 were entered after the subject's payoff regressors and regressors were orthogonalized in ascending order. This means that in case of shared variance between these regressors, all commonly explained BOLD variance was attributed to the subject's regressors, yielding an independent and conservative estimate for the effect of the receiver's payoff regressors.

In order to test our hypotheses we contrasted the following decision types to describe the neuronal correlates of social decision making during event 2. We contrasted social decisions in NCS- and E-condition with self-interested decision in the PSI-condition (NCS > PSI, E > PSI). For this decision-related activity, we build differential t-contrasts on first level. Prosocial decisions in the CS-condition (CS+) were too rare to have reliable parameter estimates. We subjected the regressors of the parametric modulators of event 1 and event 3 to one-sample *t*-tests. We tested all contrasts in the whole sample ($n = 36$) and for group differences (prosocial versus selfish participants).

The voxel-level threshold was set to 0.005 (uncorrected). We applied a whole-brain cluster-level family-wise error (FWE) correction for multiple comparisons with a cluster-p-value of 0.05. Additionally we applied a small-volume correction for a-priori defined Regions of Interest (ROI).

ROI DEFINITION

We defined regions known to be involved in valuation and reward processing as regions of interest, namely the Nucleus Accumbens (NAcc) and the medial orbitofrontal cortex (mOFC). Additionally, we were interested in the subgenual ACC, as this region was shown to be involved in reward-processing in a social context (Moll et al., 2006; Kuss et al., 2013). We derived anatomical masks of these regions from the Harvard-Oxford cortical and subcortical structural atlases (http://www.cma.mgh.harvard.edu), applying a probability of 0.5. Further we used the AAL as implemented in SPM 8 to derive masks for Caudate and Putamen as approximation to a mask for the Ventral Striatum.

RESULTS

BEHAVIORAL RESULTS

Decisions in conditions without conflict (non-costly social and efficiency) demonstrate that individuals make social decisions that profit the other person at no cost to self-interest (**Table 1**). In pure self-interest (PSI), non-costly social (NCS) and efficiency (E) situations, all subjects consistently chose the advantageous alternative (>92% of all trials over all subjects): the choices in PSI were advantageous with respect to self-interest, in NCS advantageous with respect to social gain, and in E advantageous with respect to both. **Table 1** shows the choice behavior in the four conditions separately for prosocial and selfish individuals.

In costly social situations which were characterized by a conflict between prosocial and self-interest motives, prosocial participants were more often willing to forgo own monetary advantages in order to distribute more money to the receiver. Prosocials more often chose the prosocial alternative in this condition when compared to selfish participants. For the other conditions there was no significant difference in choice behavior between the groups (see **Table 1**).

In the analysis of reaction-times we considered decisions that were also included in the GLM of the fMRI-analysis. Reaction-time refers to the event where subjects choose one alternative (**Figure 1B**) and is defined as time from appearance of the second alternative (**Figure 1B**) until button press

(choosing one alternative).There was a significant difference in reaction-times between conditions [$F_{(3,\ 102)} = 41.13$, $p < 0.001$] and an interaction-effect [condition X group: $F_{(3,\ 102)} = 14.59$, $p < 0.001$]. Reaction-times are shown in **Figure 2**, separately for prosocials and selfish participants. Prosocials took longer to decide in every experimental condition compared to selfish individuals. The group-differences were significant in PSI-condition [$t_{(34)} = 2.21$, $p = 0.034$] and for egoistic decisions in CS-condition [$t_{(34)} = 3.29$, $p = 0.002$]. The group-difference reaches a trend for choices in the E-condition [$t_{(34)} = 1.74$, $p = 0.09$], while there was no significant difference observed in the NCS-condition [$t_{(34)} = 0.52$, $p = 0.609$]. The pattern of reaction-time-differences is remarkable: Prosocials took longest for egoistic choices in the costly-social condition, whereas selfish participants took longest for social choices in the non-costly-social condition.

fMRI-RESULTS

The aim of the study was to describe differences in neuronal correlates of social-decision making between selfish- and prosocial-oriented subjects. After reporting results in the whole sample ($n = 36$), we present the group-difference between prosocial ($n = 20$) and selfish participants ($n = 16$) for the contrast of main interest (NCS > PSI).

Non-costly social decisions (NCS) compared to self-interested choices (PSI) were associated with activations in two clusters located in medial frontal regions. One cluster was located in ventromedial areas, including the medial orbitofrontal cortex (from now on: ventromedial prefrontal cortex, vmPFC). The other cluster is more dorsal, located in BA9 (from now on: dorsomedial prefrontal cortex, dmPFC). Those two clusters survive correction for multiple comparison on a whole brain level: Non-costly social decisions (NCS) were associated with a stronger BOLD-Signal in the vmPFC [NCS > PSI: MNI-coordinates of peak voxel: $X = 0$, $Y = 35$, $Z = 4$, $t = 5.53$, $k = 340$, pFWE(whole brain cluster level)<0.05] and in the dmPFC [NCS > PSI: MNI-coordinates of peak voxel: $X = -9$, $Y = 56$, $Z = 34$, $t = 4.15$, $k = 204$, pFWE(whole brain cluster level)<0.05], as shown in **Figure 3**.

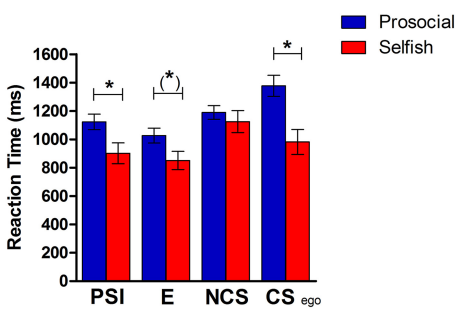

FIGURE 2 | Reaction-times in the four conditions separately for the group of prosocial and selfish participants. PSI, choosing the self-interest alternative in pure self-interest condition; E, choosing efficient alternative in efficiency condition; NCS, choosing social alternative in the non-costly social condition; CS ego, choosing the self-interest alternative in costly-social condition. *$p < 0.05$; (*)$p < 0.1$.

For the opposite contrasts (PSI > NCS), there was no activation in reward-related areas that survived correction for multiple comparisons.

These correlates of social decision making (NCS > PSI) were tested for group-differences between prosocial and selfish participants. There was no stronger activity in the group of prosocials (prosocials > proselfs). Instead there was a stronger BOLD-signal in the group of proselfs for social compared to self-interested choices (NCS > PSI) in the dmPFC [MNI-coordinates of peak voxel: $X = 0$, $Y = 32$, $Z = 34$, $t = 3.86$, $k = 172$, pFWE(whole brain cluster level)<0.05] and in the mOFC [MNI-coordinates of peak voxel: $X = 6$, $Y = 47$, $Z = -14$, $t = 5.01$, pFWE(small-volume corrected)<0.05] when compared with the group of prosocials (proselfs > prosocials, see **Figure 4**).

Further results of prosocial decision making in the efficient condition and results of the RPE-event are reported in the Supplementary Material.

DISCUSSION

This study demonstrates inter-individual differences in behavioral and neuronal correlates of prosocial preferences between prosocial and selfish individuals (Van Lange, 1999). We found that prosocials more frequently choose to allocate money to the receiver despite own losses in the costly social condition upon comparison to selfish pariticipants. This finding is more or less circular because the SVO measure used to discriminate the groups actually uses a very similar task. In fact, this result implies a validation for the fact that our fMRI paradigm resembles the definition of the SVO concept.

For the purpose of this paper, we focused on a condition where subjects could make prosocial choices (administering more money to the receiver) without any consequences for the own

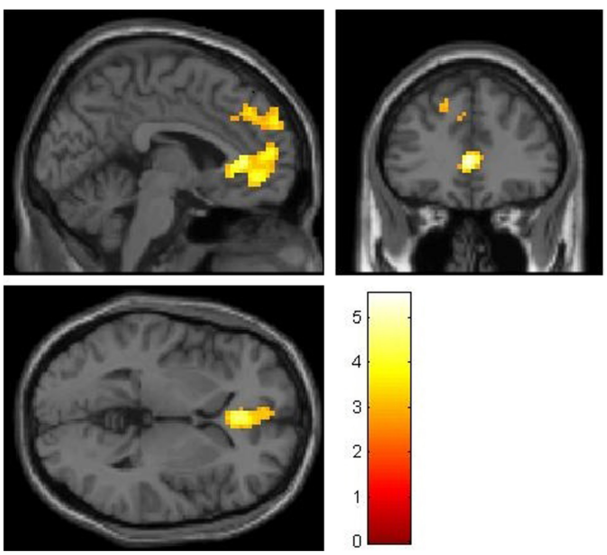

FIGURE 3 | Stronger BOLD-signal for non-costly social choices compared to pure-self interested choices (NCS > PSI) in vmPFC and dmPFC in the whole sample ($n = 36$). MNI: $X = -3$, $Y = 35$, $Z = 2$, thresholded at $t > 2.73$, corresponding to $p < 0.005$.

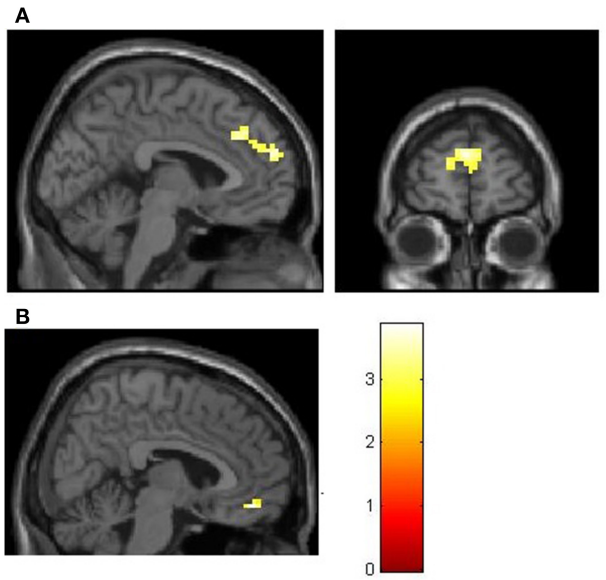

FIGURE 4 | Stronger BOLD-signal in the group of selfish participants for social compared to self-interested choices (NCS > PSI) in the dmPFC (A) and mOFC (B). **(A)** MNI: $X = -3$, $Z = 22$. **(B)** MNI: $X = 6$. Both are thresholded at $t > 2.73$, corresponding to $p < 0.005$.

payoff (non-costly social condition). In this condition the large majority of subjects chose the prosocial alternative and selfish and prosocial participants did not differ in this issue. Crucially we investigated reaction time differences and differences in BOLD signal during those choices in order to elucidate the underlying cognitive processes and their differences in the two groups.

Reaction times were significantly longer in prosocial than in selfish individuals in all conditions except for the non-costly social (NCS) condition yielding a significant group x condition interaction. A similar interaction was found for BOLD activity in the vmPFC and dmPFC, where the contrast between NCS and PSI (pure self-interest) was larger for selfish than for prosocial individuals. The NCS condition is the only condition where self-beneficial choices cannot be made. Selfish individuals are therefore required to overcome their default of primarily considering their own outcome.

GROUP DIFFERENCES IN REACTION TIMES AS INDICATORS OF AN EGOISTIC DEFAULT IN SELFISH INDIVIDUALS

One possibility to experimentally disentangle automated and controlled processes is to manipulate cognitive load (Satpute and Lieberman, 2006). Cornelissen et al. (2011) studied the effects of social value orientation on prosocial behavior in a dictator-game under cognitive load during which prosocials transferred more money to the recipient compared to selfish participants. The authors concluded that chronically accessible values are automatically transferred into behavior.

In our paradigm, group differences in reaction-times allow for conclusions on the automaticity of decision-processes in prosocial and selfish participants. Selfish participants react faster in all conditions compared to prosocials. Those differences were significant, except for non-costly social decisions. We assume that selfish participants automatically chose the options with the higher payoff for themselves. This option is easy to determine in all conditions, except for the non-costly social condition (here the own payoff is constant in both alternatives). In this condition, selfish-oriented participants are prompted to look at the payoff for the receiver and take it into account. Prosocial-oriented participants instead consider the receiver's payoff in all decisions, thus taking longer to decide in general.

This is clearly the case in the costly social condition: Here prosocial participants take very long to decide, which reflects the conflict between self-interest and prosocial motives in this condition. Selfish participants instead are quite fast in this decision, presumably because they use the heuristic of choosing the alternative with the highest payoff for themselves. While the fast reactions of selfish participants can be regarded as an indicator of intuition in following an egoistic decision default, the slow responses of prosocials can be seen as an indicator of deliberation in the costly social condition where self-interest and prosocial motives are in conflict (Rubinstein, 2007; Kahneman, 2011). In the underlying value-based decision process, prosocials obviously place a positive value on the outcome of the other person, whereas selfish indiviuals primarily value their own outcomes.

NEURAL CORRELATES OF OVERRIDING THIS EGOISTIC DEFAULT IN AREAS OF VALUE COMPUTATION, COGNITIVE CONTROL AND SOCIAL COGNITION ARE MORE PRONOUNCED IN SELFISH INDIVIDUALS

Our results demonstrate activity during social choices in brain areas that are associated with on the one hand (1) subjective (reward-) value (mOFC, vmPFC) and on the other hand (2) theory of mind, executive function and cognitive control (dmPFC).

Besides its reward-related function, the vmPFC is also associated with the integration of costs and benefits (De Quervain et al., 2004; Basten et al., 2010) and with choosing alternatives with high subjective value (Rangel and Hare, 2010; Bartra et al., 2013). Decisions in our paradigm require the integration of the value of the personal payoff and the value of the receiver's payoff into one subjective value. The vmPFC activity in our paradigm was observed for non-costly social decisions. Non-costly social choices in our paradigm induce selfish individuals to consider the payoff of another person, and thus induces a valuation-process that adds on the selfish default of considering mainly their own payoff. The vmPFC activity was stronger in the group of selfish participants implying higher demand on value computation in selfish individuals, compared to the more intuitive prosocial decisions of prosocials.

In a similar vein, higher activity in the dmPFC in selfish individuals during non-costly social choices can be seen as a correlate of reflection and displays the need of cognitive resources (and thus less automaticity) in individuals who primarily use an egoistic decision default.

The dmPFC is associated with cognitive control and controlled forms of social cognition as opposed to automatic forms of social cognition (Satpute and Lieberman, 2006; Lieberman, 2007). Different cognitive processes have been reported to be associated with dmPFC activity, especially for tasks that require

cognitive control and computational load (e.g., Elliott and Dolan (1998) for higher cognitive demands during hypothesis testing; Ferstl and von Cramon (2002) during coherency processing of speech; Berthoz et al. (2002) during processing of norm violations; Decety et al. (2004) during competition). Those results demonstrate the dmPFC' s role for the maintenance of non-automated cognitive processes and hint at a more general and domain-independent function (Ferstl and von Cramon, 2002). In this vein, the dmPFC activity during non-costly social choices can be regarded as an indicator of a reflective cognitive process, which is stronger in selfish-oriented participants.

The dmPFC (BA9) is also associated with the processing of socially relevant stimuli, theory of mind and mentalizing (Gallagher and Frith, 2003; Saxe, 2006). Stronger activation during social decisions in selfish participants could also reflect a higher demand of theory of mind- and executive-functions in this group. In a similar vein, Krueger et al. (2007) report group-differences in BA 9 during the course of a trust-game: the group of participants that did not experience reciprocity in the first half of the experiment had stronger activity in BA9 compared to participants who experienced their trust being reciprocated. Participants that did not experience reciprocity rely more on mentalizing processes in order to predict the behavior of the other. Participants that did however experience reciprocity during the game, trust more automatically and thus demand fewer mentalizing processes. (Krueger et al., 2007).This parallels our group-differences during social choices in BA9, because we assume prosocials to use social cognition more automatically compared to selfish individuals.

Neural indicators for a reward value of prosocial decisions were observed as well. Activity in the mOFC and vmPFC during social choices imply that decisions, which profit another person at no cost to the decision maker, carry an intrinsic reward-value. Our results confirm results of previous studies that reported reward-related activity during social choices in interactive cooperation paradigms (Decety et al., 2004; Rilling et al., 2004; Elliott et al., 2006; Emonds et al., 2011), as well as in non-interactive paradigms (Moll et al., 2006; Mobbs et al., 2009; Tricomi et al., 2010; Zaki and Mitchell, 2011; Fareri et al., 2012).

To conclude, our results hint at the need for controlled processes in prosocial decision making. Additionally prosocial decisions at no cost to self-interest seem to have an intrinsic reward-value. Furthermore our results demonstrate interindividual differences in neuronal correlates of prosocial decision making. These decisions seem to recruit the need to overcome the default of maximizing self-interest in selfish individuals, which is accompanied by activity in areas associated with controlled forms of social cognition such as the dmPFC (Lieberman, 2007) and areas associated with the integration of values such as the vmPFC (Rangel and Hare, 2010). The results allow a more detailed view on prosocial decision making that takes interindividual differences into account: it does not seem to be a question of deliberation versus intuition *per se*. Instead, selfish-oriented individuals apply reflection and need cognitive resources to overcome self-interest and act prosocially, whereas prosocial-oriented individuals seem to rely more on intuitive processes.

AUTHOR CONTRIBUTIONS

KK, KF, BW, CM, AF designed the experiments. KK, CM, PT programmed the experiment. KK and CM conducted the experiment. KK, PT, KF analyzed data. BW, AF, KF reviewed and supervised data analysis. KK, KF, CM, BW, AF discussed and interpreted the data. KK, KF, CM, BW wrote the manuscript.

ACKNOWLEDGMENTS

KK and KF were funded by the German Research Council (Grant FL 715/1-1). BW is supported by a Heisenberg Grant of the German Research Council (Grant We 4427/3-1). CM is funded by a Heisenberg grant of the German Research Foundation (Grant MO-2363/3-1).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fnbeh.2015.00040/abstract

REFERENCES

Bartra, O., McGuire, J. T., and Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76, 412–427. doi: 10.1016/j.neuroimage.2013.02.063

Basten, U., Biele, G., Heekeren, H. R., and Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21767–21772. doi: 10.1073/pnas.0908104107

Berthoz, S., Armony, J. L., Blair, R. J. R., and Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain* 125, 1696–1708. doi: 10.1093/brain/awf190

Bogaert, S., Boone, C., and Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: a review and conceptual model. *Br. J. Soc. Psychol.* 47, 453–480. doi: 10.1348/014466607X244970

Boone, C., Declerck, C., and Kiyonari, T. (2010). Inducing cooperative behavior among proselfs versus prosocials: the moderating role of incentives and trust. *J. Confl. Resolut.* 54, 799–824. doi: 10.1177/00220027 I 0372329

Boone, C., Declerck, C., and Suetens, S. (2008). Subtle social cues, explicit incentives, and cooperation in social dilemmas. *Evol. Hum. Behav.* 29, 179–188. doi: 10.1177/00220027 I 0372329

Camerer, C. F. (2003).*Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ:Princeton University Press.

Cornelissen, G., Dewitte, S., and Warlop, L. (2011). Are social value orientations expressed automatically? Decision making in the dictator game. *Pers. Soc. Psychol. Bull.* 37, 1080–1090. doi: 10.1177/0146167211405996

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., and Meltzoff, A. N. (2004). The neural bases of cooperation and competition: an fMRI investigation. *Neuroimage* 23, 744–751. doi: 10.1016/j.neuroimage.2004.05.025

Declerck, C. H., Boone, C., and Emonds, G. (2013). When do people cooperate? The neuroeconomics of prosocial decision making. *Brain Cogn.* 81, 95–117. doi: 10.1016/j.bandc.2012.09.009

De Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., et al. (2004). The neural basis of altruistic punishment. *Science* 305, 1254–1258. doi: 10.1126/science.1100735

Elliott, R., and Dolan, R. J. (1998). Activation of different anterior cingulate foci in association with hypothesis testing and response selection. *Neuroimage* 8, 17–29. doi: 10.1006/nimg.1998.0344

Elliott, R., Völlm, B., Drury, A., McKie, S., Richardson, P., and Deakin, J. F. W. (2006). Co-operation with another player in a financially rewarded guessing game activates regions implicated in theory of mind. *Soc. Neurosci.* 1, 385–395. doi: 10.1080/17470910601041358

Emonds, G., Declerck, C. H., Boone, C., Vandervliet, E. J. M., and Parizel, P. M. (2011). Comparing the neural basis of decision making in social dilemmas of people with different social value orientations, a fMRI study. *J. Neurosci. Psychol. Econ.* 4, 11–24. doi: 10.1037/a0020151

Engel, C. (2011). Dictator games: a meta study. *Exp. Econ.* 14, 583–610. doi: 10.1007/s10683-011-9283-7

Fareri, D. S., Niznikiewicz, M. A., Lee, V. K., and Delgado, M. R. (2012). Social network modulation of reward-related signals. *J. Neurosci.* 32, 9045–9052. doi: 10.1523/JNEUROSCI.0610-12.2012

Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* 11, 419–427. doi: 10.1016/j.tics.2007.09.002

Ferstl, E. C., and von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: coherence or theory of mind? *Neuroimage* 17, 1599–1612. doi: 10.1006/nimg.2002.1247

Gallagher, H. L., and Frith, C. D. (2003). Functional imaging of "theory of mind." *Trends Cogn. Sci.* 7, 77–83. doi: 10.1016/S1364-6613(02)00025-6

Haruno, M., and Frith, C. D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nat. Neurosci.* 13, 160–161. doi: 10.1038/nn.2468

Haruno, M., Kimura, M., and Frith, C. D. (2014). Activity in the nucleus accumbens and amygdala underlies individual differences in prosocial and individualistic economic choices. *J. Cogn. Neurosci.* 26, 1861–1870. doi: 10.1162/jocn_a_00589

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832. doi: 10.1126/science.1129156

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., et al. (2007). Neural correlates of trust. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20084–20089. doi: 10.1073/pnas.0710103104

Kuss, K., Falk, A., Trautner, P., Elger, C. E., Weber, B., and Fliessbach, K. (2013). A reward prediction error for charitable donations reveals outcome orientation of donators. *Soc. Cogn. Affect. Neurosci.* 8, 216–223. doi: 10.1093/scan/nsr088

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.* 58, 259–289. doi: 10.1146/annurev.psych.58.110405.085654

Mobbs, D., Yu, R., Meyer, M., Passamonti, L., Seymour, B., Calder, A. J., et al. (2009). A key role for similarity in vicarious reward. *Science* 324, 900. doi: 10.1126/science.1170539

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., and Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15623–15628. doi: 10.1073/pnas.0604475103

Piovesan, M., and Wengström, E. (2009). Fast or fair? A study of response times. *Econ. Lett.* 105, 193–196. doi: 10.1016/j.rconlet.2009.07.017

Rand, D. G., Greene, J. D., and Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature* 489, 427–430. doi: 10.1038/nature11467

Rangel, A., and Hare, T. (2010). Neural computations associated with goal-directed choice. *Curr. Opin. Neurobiol.* 20, 262–270. doi: 10.1016/j.conb.2010.03.001

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., and Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 22, 1694–1703. doi: 10.1016/j.neuroimage.2004.04.015

Rubinstein, A. (2007). Instinctive and cognitive reasoning: a study of response times. *Econ. J.* 117, 1243–1259. doi: 10.1111/j.1468-0297.2007.02081.x

Satpute, A. B., and Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Res.* 1079, 86–97. doi: 10.1016/j.brainres.2006.01.005

Saxe, R. (2006). Uniquely human social cognition. *Curr. Opin. Neurobiol.* 16, 235–239. doi: 10.1016/j.conb.2006.03.001

Tricomi, E., Rangel, A., Camerer, C. F., and O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature* 463, 1089–1091. doi: 10.1038/nature08785

Van den Bos, W., van Dijk, E., Westenberg, M., Rombouts, S. A., and Crone, E. A. (2009). What motivates repayment? Neural correlates of reciprocity in the Trust Game. *Soc. Cogn. Affect. Neurosci.* 4, 294–304. doi: 10.1093/scan/nsp009

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: an integrative model of social value orientation. *J. Pers. Soc. Psychol.* 77, 337–349.

Zaki, J., and Mitchell, J. P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19761–19766. doi: 10.1073/pnas.1112324108

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 November 2014; accepted: 05 February 2015; published online: 24 February 2015.

Citation: Kuss K, Falk A, Trautner P, Montag C, Weber B and Fliessbach K (2015) Neuronal correlates of social decision making are influenced by social value orientation—an fMRI study. Front. Behav. Neurosci. ***9****:40. doi: 10.3389/fnbeh.2015.00040*

This article was submitted to the journal Frontiers in Behavioral Neuroscience.

doi:10.1093/scan/nsr088 SCAN (2013) 8, 216–223

A reward prediction error for charitable donations reveals outcome orientation of donators

Katarina Kuss,[1,2] Armin Falk,[2] Peter Trautner,[3] Christian E. Elger,[1,2,3] Bernd Weber,[1,2,3] and Klaus Fliessbach[1,2,3]

[1]Department of Epileptology, University Hospital Bonn, Sigmund-Freud-Str. 25, [2]Center for Economics and Neuroscience, University of Bonn, Nachtigallenweg 86 and [3]Life & Brain Center, Department of NeuroCognition, University of Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany

The motives underlying prosocial behavior, like charitable donations, can be related either to actions or to outcomes. To address the neural basis of outcome orientation in charitable giving, we asked 33 subjects to make choices affecting their own payoffs and payoffs to a charity organization, while being scanned by functional magnetic resonance imaging (fMRI). We experimentally induced a reward prediction error (RPE) by subsequently discarding some of the chosen outcomes. Co-localized to a nucleus accumbens BOLD signal corresponding to the RPE for the subject's own payoff, we observed an equivalent RPE signal for the charity's payoff in those subjects who were willing to donate. This unique demonstration of a neuronal RPE signal for outcomes exclusively affecting unrelated others indicates common brain processes during outcome evaluation for selfish, individual and nonselfish, social rewards and strongly suggests the effectiveness of outcome-oriented motives in charitable giving.

Keywords: charitable donations; reward processing; social preferences; functional magnetic resonance imaging (fMRI); nucleus accumbens (NAc)

INTRODUCTION

There is compelling evidence that humans do not exclusively follow rational, self-interested motives in economic decision making (Camerer and Fehr, 2006). Perhaps the most striking exception to materialistic, self-interested behavior is giving one's own goods to unrelated others, as is in charitable donations.

Theoretical underpinnings of donation behavior

Economic theory suggests that different motives underlie donation decisions (Harbaugh, 1998). One possible motive is that a person has in fact a preference for the public good provided by donations. In this case, the money belonging to a charity organization carries a utility (or reward value) that is independent of the person's own contribution. Such a preference for a public good can be regarded as 'altruistic' and if this is the sole motivation for a donation, this behavior has been termed 'pure altruism' (Andreoni, 1989). Notably, pure altruism does not imply a noninterest in own belongings. According to Andreoni, pure altruism means that 'preferences depend only on private consumption and the total supply of the public good . . .' (Andreoni, 1990). Because the preference for the public good is directed toward the outcome of a donation, we refer to this motive as 'outcome orientation'.

Apart from this, it has been argued that other motives, such as guilt avoidance, reputation gain or a feeling of 'warm glow', can be associated with *the act of giving itself* ('action orientation'; Andreoni, 1989, 1990). Action- and outcome-oriented motives do not contradict each other. Rather, they are supposed to complement each other, empirical support for the effectiveness of both motives in charitable donations comes from behavioral studies (Andreoni, 1989, 1990; Konow, 2010). Andreoni refers to this as 'impure altruism'.

Neuroimaging findings in donation behavior

Recently, functional neuroimaging has been used to investigate brain processes underlying donation behavior. These studies have shown that donation decisions are associated with activations in the dopaminergic reward system (Moll *et al.*, 2006; Harbaugh *et al.*, 2007), providing support for action-associated positive feelings in the sense of a warm glow. At the same time, indirect support for outcome orientation has been found by showing increased reward-related brain activity during nonvoluntary transfers of money to a charity that can be used to predict subjects' donation behavior (Harbaugh *et al.*, 2007).

Reward prediction error induction as a test of outcome orientation

Our study aims at extending these findings by directly probing outcome-related reward activity in the context of

Received 22 July 2011; Accepted 14 November 2011
Advance Access publication 23 December 2011

K.K. and K.F. are funded by the German Research Council (Grant FL 715/1-1). B.W. is funded by the German Research Council with a Heisenberg Grant (Grant WE 4427/3-1). The authors thank Florian Mormann and Johannes Niediek for helpful comments on the manuscript.

Correspondence should be addressed to Klaus Fliessbach, MD, Department of Epileptology, University of Bonn Medical Center, Sigmund Freud-Str. 25, D-53127 Bonn, Germany.
E-mail: klaus.fliessbach@ukb.uni-bonn.de

charitable donations. For this purpose, we designed an experimental situation in which subjects make decisions about the allocation of money to themselves and to a charity. After their decisions have been made, some of these decisions are discarded while others are confirmed, which allows the definition of two different reward prediction errors (RPEs): one with respect to the own, personal payoff and one with respect to the charity's payoff. The term RPE originates from reinforcement learning, where RPEs are assumed to drive adaptive learning (Knutson *et al.*, 2000). In a broader sense, the term has been applied to all situations in which a mismatch between expected and actual outcome occurs, even in the absence of learning, such as in guessing (Yacubian *et al.*, 2006) or lottery tasks (Breiter *et al.*, 2001). We apply the term RPE in this broader sense, i.e. RPEs arise whenever rewards are not fully predictable. After subjects make their choice in our experiment, there is uncertainty whether this choice will be subsequently confirmed or discarded. Therefore, an RPE arises at the time when subjects are informed about the confirmation or discard of their choice.

Questions and hypotheses

At the neural level, RPEs for one's own material goods are represented in the dopaminergic mesolimbic system (Schultz, 1998), and human fMRI studies reliably detect a corresponding BOLD signal in the nucleus accumbens (NAc; Pagnoni *et al.*, 2002). Therefore, in our study, we expected to detect a signal corresponding to the RPE for personal payoffs in the NAc. The key question was whether we would observe an equivalent signal for payoffs concerning the charity organization. We expected such a signal in subjects that made donations at their own costs. If outcome orientation has motivated these subjects' donation decisions, they should attribute a reward value to the charity's payoff, and consequently our RPE manipulation should be associated with an RPE signal similar to that for personal rewards. In addition to the NAc, we tested for outcome-related activity in other brain areas relevant to reward processing in a prosocial context, such as the subgenual area (Moll *et al.*, 2006), the medial orbitofrontal cortex (mOFC; Hare *et al.*, 2010) and the ventral tegmental area (VTA) of the midbrain (Krueger *et al.*, 2007). Furthermore, we tested whether at the time of decision making we could detect activation in any of the areas of interest that would further support the idea of action orientation.

METHODS

Participants and procedure

Prior to the fMRI experiment, each of our 33 subjects (17 female, mean age = 25.6 years, range: 21–35 years) chose one charity from a list of six organizations (Supplementary Table S1) that would benefit from her decisions. In each of the 180 trials of the fMRI experiment, subjects chose one of two alternatives, each consisting of a payoff for themselves and for the charity (Figure 1). Subjects were informed that the payoff of one randomly chosen trial would be implemented after the experiment (*actual payoff*). There was no deception: the selection of the implemented trial was random and the *actual payoffs* were transferred to the subjects and the charities. Subjects were guaranteed anonymity of their decisions (see Methodological Details in Supplementary Data).

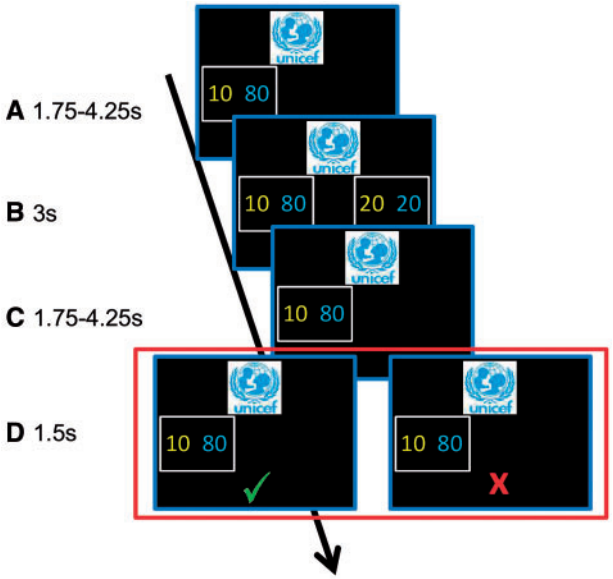

Fig. 1 Single trial settings. On the first screen, subjects saw the first alternative comprising of a payoff for the subject in yellow and for the charity in blue (A). After a jittered time interval, the second alternative appeared. Subjects had up to 3 s to select one alternative by button-press (B). Note that the trials are randomly drawn from all possible combinations of decision situations. This results in a randomization of the order of payoff alternatives (see Methodological Details in Supplementary Material). The selected alternative was presented as a response feedback (C). After a jittered time interval, a fourth screen appeared, informing subjects whether the trial was discarded (red cross) or confirmed (green check) to be among the trials from which the actual payoff trial would be chosen (D). In the second part of the experiment, confirmed outcomes from part 1 were presented (screen one in part 2; C in Figure 1), subjects were informed on a second screen whether the trial was discarded or confirmed (D).

Immediately after their decisions in each trial, subjects were informed whether this trial would be considered for the selection of the *actual payoff* (first RPE induction). At this point, 50% of the trials were randomly discarded. After the decision experiment, a second fMRI session followed, during which subjects saw each of the 90 chosen alternatives that had been confirmed during the first session. Fifty percent of these alternatives were again discarded (second RPE induction). At the time of the first RPE induction, reactions to the presented outcome might still be influenced by the previous decision. In contrast, the outcomes presented in the second session cannot be related to the decision they were based on. We therefore regard the point of the second RPE induction as the one that best represents pure outcome evaluation.

Subject's and charity's payoffs varied independently among 5, 10, 20, 40 and 80€, and payoff alternatives were

randomly chosen from all possible unique combinations of different alternatives (Supplementary Figure S1). This led to four qualitatively distinguishable decision situations: 'pure self-interest' (PSI), 'noncostly donation' (NCD), 'efficiency' (E) and 'costly donation' (CD) situations. In the costly donation situations, subjects could choose to forgo monetary advantages to allocate more money to the charity (e.g. favoring 10€ for themselves and 80€ for the charity over 20€ for both; see Figure 1). In this situation, subjects had to trade material self-interest with altruistic preferences. The other three situations do not entail a conflict between different motives because one alternative is unequivocally advantageous with respect to self-interest motives (PSI), efficiency (NCD) or both motives (E). These different decision situations parallel to those implemented in a study by Moll *et al.* (2006). In contrast to Moll *et al.*, the subjects in our experiment chose between two payoff alternatives instead of having a choice to accept or reject a single payoff distribution. This allowed us to test whether at the time of presentation of the first alternative, there was a brain signal predicting the upcoming decision in the sense of a value computation. However, we did not observe such a signal (not reported). The most important extension in comparison to the study by Moll *et al.* (2006) consists of the RPE induction manipulations, which allow us to observe outcome-related effects without confounding decision-related effects.

fMRI analysis

General linear model

We included three events for the first part of the experiment: onset of the appearance of the first alternative (event 1), different onset-regressors, depending on the decision situations (event 2), onset of RPE-induction, including two parametric modulators representing the RPE of the subject's payoff, and the RPE of the charity's payoff (event 3). For the second part of the experiment, two events were included: the appearance of the chosen alternative (event 4) and the onset of the second RPE induction (event 5), including the two parametric RPE regressors (Supplementary Table S2). The parametric regressors for the RPEs were collapsed over all decision types; the rationale for this was to test whether there is an RPE signal for an outcome which is independent of the decision it is based on. This is based on the assumption that the money belonging to the charity organization carries a reward value for subjects with prosocial preferences independent of their own contribution (see 'Introduction' section).

Importantly, the parametric regressor for the charity's payoff RPE was entered after the subject's payoff RPE regressor and regressors were orthogonalized in ascending order. This means that in case of shared variance between these regressors, all commonly explained BOLD variance was attributed to the subject's payoff RPE regressor, yielding an independent and conservative estimate for the effect of the charity's payoff RPE (for Preprocessing and further information, see Details of Analysis in Supplementary Data).

Reward prediction error model

The RPE is defined as the difference between the reward magnitude (RM) of an outcome and the expected value (EV). In our experiment, the RM is not simply reflected by the respective absolute payoff of a choice (x), but must be computed with respect to a reference point (RP): $RM = x–RP$. The RP depends on the subjects' previous experiences from the experiment through facing the range of possible payoffs. We assumed that subjects attribute a negative RM to payoffs that are lower than their RP. We defined the RP as the median of all previous experienced payoffs, since this resulted in the strongest effects for the subject's payoff RPE. (We also tested alternative RPs and resulting RPEs. For a discussion, see Details of Analysis in Supplementary Data and Supplementary Table S6). Thus RPE is defined as RM–EV when a choice was confirmed, and 0–EV when a choice was discarded, respectively, with $EV = 0.5 \times RM$, given a 50% chance that a choice is confirmed or discarded.

Region of interest definition

We focused the analysis on the NAc and derived anatomical masks of this region from the Harvard–Oxford cortical and subcortical structural atlases (http://www.cma.mgh.harvard.edu), applying a probability of 0.5. In the same way, we generated anatomical masks for the subgenual area and the mOFC. Note, that these two ROIs are not fully covered by the EPI images (Supplementary Figure S2). To cover the VTA, we used an anatomical mask of the entire midbrain posteriorly cut off at MNI coordinate $y = -22$. For the analysis of the RPE induction in the NAc, we extracted parameter estimates for the subjects' and the charity's payoff RPE from the NAc masks, averaged across all voxels. For the other regions, small volume corrections for multiple comparisons [family wise error (FWE)] were applied because we did not expect an average effect across these relatively large and functionally heterogeneous areas. In addition, we performed a whole brain conjunction analysis to test for overlapping effects of the subject's payoff RPE and the charity's payoff RPE regressor in the donator group [minimum statistic against conjunction null at $P < 0.001$, uncorrected for each individual contrast, see Nichols *et al.* (2005)].

RESULTS

Behavioral results

In situations PSI, NCD and E, all subjects consistently chose the advantageous alternative (>98% of all trials over all subjects; Table 1). We classified the few instances of deviant choices as implausible decisions and assumed that they were based on accidental errors. High inter-individual variance was only observed for the CD situation, with 17.4 % (s.d. = 25.17%) of all trials over all subjects being costly

Table 1 The four decision situations and their underlying payoff-structures (A1/B1 A2/B2) including percentages of subjects choosing A1/B1 in each situation

Decision situation	Payoff structure	Percentage of subjects choosing A1/B1 (mean ± SD)
Pure self-interest (PSI)	A1>A2, B1 = B2 e.g. 10/20 5/20	98.64 (±2.49)
Noncostly donation (NCD)	B1>B2, A1 = A2 e.g. 5/40 5/10	97.9 (±3.22)
Efficiency (E)	A1>A2, B1>B2 e.g. 20/40 10/20	99.48 (±0.93)
Costly donation (CD)	A1<A2, B1>B2 e.g. 5/80 20/40	17.49 (±25.17)

Notes: A: subject's payoff, B: charity's payoff, A1/B1: first alternative, A2/B2: second alternative. Note that the trials are randomly drawn from all possible combinations of decision situations. This results in a randomization of the order of payoff alternatives (see Methodological Details in Supplementary Data).

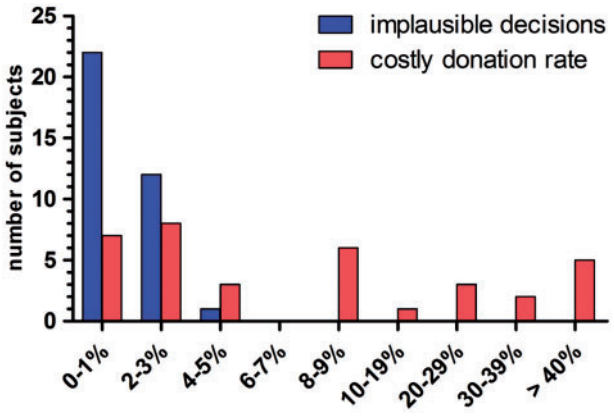

Fig. 2 Donation behavior. Distribution of subjects with respect to donation rates (CD^+) and rates of implausible decisions; CD^+: choosing donation-alternative in CD situations.

donation decisions. To confidently identify subjects who intentionally donated money in the costly donation situation, we applied the following statistical criterion: Subjects with a donation rate (CD^+) significantly (Fisher's exact test, $P < 0.05$) higher than the same subjects' rate of implausible decisions were classified as 'donators' ($n = 16$) and the others as 'nondonators' ($n = 17$) (Figure 2). Support for this separation comes from the fact that donators engaged more frequently in real-life prosocial activities than nondonators, according to a self-report questionnaire (Additional Results in Supplementary Data). The rates of costly donations within the donator group ranged from 8% (5 out of 60) to 97% with a mean of 33.7%. There were significant reaction time (RT) differences between the decision situations [main-effect of decision situation: $F(3, 93) = 44.848$, $P > 0.001$]. In the subgroup of donators RT in costly donation situations (CD^+) were longer than in noncostly donations (NCD^+), pure self-interest situations (PSI^+) and efficiency situations (E^+). There was also a significant decision situation × subgroup interaction $F(3, 93) = 16.686$, $P < 0.001$. *Post hoc* *t*-tests reveal faster reactions of the nondonators in pure self-interest situations (PSI^+), and for self-interest decisions in costly donation situations (CD^-, Supplementary Table S4). Donators and nondonators did not differ in demographic variables, such as age, monthly income or gender (Supplementary Table S3).

fMRI results

fMRI data analysis focused on the two RPE induction events. Further, we tested whether different decision types revealed different activation levels in the NAc at the time of decision making. We did not find any association: irrespective of the decision type, there was a positive BOLD signal in the NAc during decision making compared to unmodeled baseline activity (Supplementary Figures S3 and S4). Further, there were no significant differences between donation decision trials and any other conditions in any of the ROIs or between donators and nondonators (reported in Supplementary Data). These results do not support the existence of differential decision-related reward system activity for donation decisions.

For the outcome phase, BOLD activity in the NAc was highly correlated with the subject's payoff RPE at the first and second RPE induction for the entire group of subjects. This sets the stage for the main question of whether we can detect a co-localized, equivalent signal for the charity's payoff RPE. Figure 3 shows our main result (averaged over both RPE induction time-points and all voxels in the NAc masks): in the NAc, there was indeed a highly significant positive modulation of BOLD activity by the charity's payoff RPE in donators [$t(14) = 4.644$, $P = 0.00018$, one-tailed] but not in the nondonators [$t(16) = 1.195$, $P = 0.125$, one tailed]. The group difference between donators and nondonators was significant [$t(30) = 2.164$, $P = 0.02$, one tailed]. The relation between RPE signal and donation behavior was further corroborated by a significant correlation between the costly donation rate as a continuous, behavioral variable and the charity's payoff RPE signal (Spearman's $\rho = 0.309$, $P = 0.043$, one tailed).

The comparison between first and the second RPE induction reveals several differences: the effect for the charity RPE regressor in donators at the first RPE induction was only significant in the right but not in the left NAc and is on average significantly weaker than at the second RPE induction (Supplementary Table S5). In addition, at the first RPE induction, the RPE signal for *personal* payoff for the nondonators was significantly lower than that of donators (Supplementary Figure S5). To further elucidate these differences, we ran an additional analysis for the first RPE induction in which all trials in the costly donation situations were excluded. The rationale for this is that we expected the evaluation of costly donation situations in donators to be

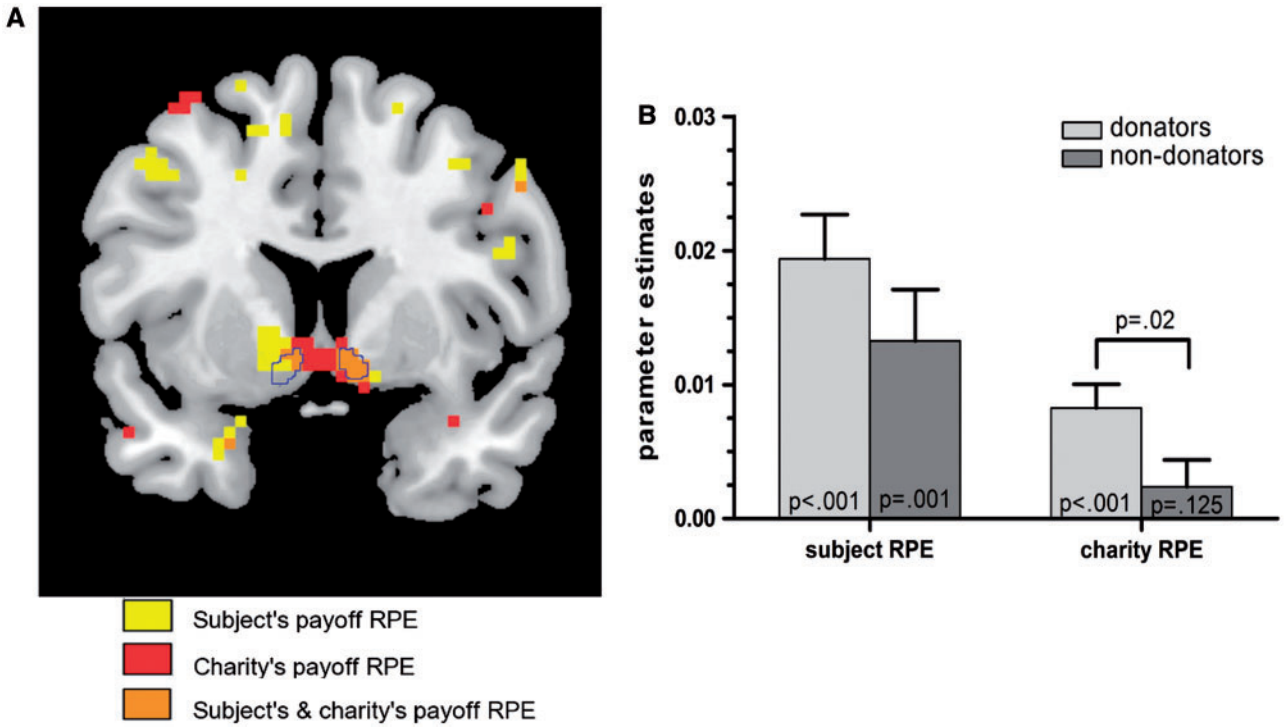

Fig. 3 FMRI results. Nucleus accumbens signals reward prediction errors for one's own payoff in all subjects and for charity's payoff in the donator group. Results are averaged across the two time points of RPE induction. (A) Coronal brain section ($y = 8$) showing voxels with a significant modulation of the BOLD signal by the subject's payoff RPE (yellow, thresholded at $t > 5.99$ corresponding to $P < 0.0001$, uncorrected) and by the charity's payoff RPE (red, $t > 2.98$, $P < 0.005$, uncorrected) for the donators ($n = 15$) (one subject did not complete part 2 of the experiment, see Methodological Details in Supplementary Material). The NAc region of interest (ROI) is framed in blue. The effect of charity's payoff RPE is significant after small-volume correction for multiple comparisons within this ROI ($P_{\text{FWE-corrected}} < 0.05$). (B) Bar plot showing mean parameter estimates (± *SEM*) for the bilateral NAc. *Ps* (one tailed) for one-sample *T*-tests against zero for each regressor and for a two-sample *T*-test are shown. For separate results at each stage of RPE induction, see Supplementary Figures S5 and S6 and Table S5.

more deliberate (reflected by longer RTs) than in the other situations which might have influenced the following outcome phase. After the exclusion of these trials, we observe a highly significant charity's payoff RPE in the donator group, which does not differ from that of the second RPE induction (Supplementary Figure S7). For separate results of the first and second RPE induction, see Supplementary Figures S5 and S6, Supplementary Table S5.

Remarkably, the activation peaks and the majority of activated voxels for the charity's payoff RPE in donators lie slightly medial outside our predefined NAc region of interest toward the septal region (activation peak at MNI coordinates: $X = -3$, $Y = 8$, $Z = -8$, $t = 6.45$, $P_{\text{uncorrected, whole brain}} < 0.001$). The activation cluster extends anteriorly into the subgenual area [MNI coordinates for peak voxel in the subgenual area ROI: $X = 3$, $Y = 11$, $Z = -8$, $t = 4.95$, $P_{\text{FWE(small-volume corrected)}} < 0.05$]. Conversely, activation peaks for the personal payoff RPE lie more laterally, within the predefined NAc masks [MNI coordinates: $X = 9$, $Y = 11$, $Z = -8$, $t = 8.79$, $P_{\text{FWE(small-volume corrected)}} < 0.05$]. For details on the spatial relations between activation clusters and ROI, see Supplementary Figures S8 and S9.

Within the other ROI, there was a significant positive modulation of BOLD activity by the subject's payoff RPE in the mOFC and ventral midbrain (because of methodological limitations of fMRI in the localization of brainstem activity we refrain from using the term VTA), but no corresponding signal for the charity's RPE (Supplementary Figures S10 and S11).

The whole-brain conjunction analysis for both regressors confirmed overlapping areas bilaterally within the NAc-ROI [$P_{\text{FWE (small-volume corrected)}} < 0.05$] (Supplementary Figure S12 and Table S7). There were no overlapping activations surviving correction for multiple comparisons (whole brain or small volume correction for other ROIs) elsewhere. Areas surviving the inclusion threshold are reported in Additional Results in Supplementary Data.

DISCUSSION

Our study investigates decision- and outcome-related reward system activity in the context of charitable donations. For this purpose, our subjects took part in an fMRI experiment in which they made decisions affecting both their personal and a charity organization's payoff. This decision experiment served two purposes: it allows us to observe brain activation accompanying donation decisions (similar to Moll *et al.*, 2006), and it allows us to identify subjects which behaviorally express their willingness to donate. The main focus,

however, was on the processing of outcomes. For this reason, we introduced a novel experimental manipulation by discarding part of the outcomes, thus introducing one RPE for the subject's personal payoff, and another RPE for the payoff to the charity organization.

In our subject group, ~50% of subjects made costly donation decisions. This is consistent with recent data from a worldwide survey that shows that 49% of the German population sometimes donates money (Charities Aid Foundation, 'The World Giving Index 2010'). Higher donation rates in other neuroimaging studies (Moll *et al.*, 2006; Harbaugh *et al.*, 2007; Hare *et al.*, 2010) might partly be explained by cultural differences. In the USA, for example, the percentage of donators in the average population is about 64%, according to the same source. Within the group of donators, there was large variability in the rate of costly donations, which is presumably due to the fact that within the costly donation situations there is large variation in the efficiency associated with each decision alternative. In many trials that involved a costly donation, subjects would have to give up more money than the charity would gain. Consequently, only one subject chose to donate in almost every case. Others only donated in situations where they had to give up little of their own to make a large donation, yielding only 5 out of 60 donation decisions. In contrast, in other neuroimaging studies (Moll *et al.*, 2006; Hare *et al.*, 2010), the payoff to the charity in costly donation decision was consistently enhanced in relation to the personal losses, which presumably promotes higher donation rates. Finally, in contrast to another neuroimaging study (Izuma *et al.*, 2010), our subjects made their decisions anonymously, so that social reputation gains were unlikely to contribute to higher donation rates.

Our main analysis addressed the question whether similar reward signals can be detected related to personal and the charity's payoff in reward processing areas of the brain; it thereby addressed the fundamental question of whether social and nonsocial cognition share common underlying brain processes (Adolphs, 2003). A number of previous studies have shown overlapping neural substrates at different stages of reward processing for nonsocial (e.g. own monetary) and social rewards (Izuma *et al.*, 2008; Zink *et al.*, 2008; Smith *et al.*, 2010; Lin *et al.*, 2011). In these studies, social rewards included rewards with a nonmonetary benefit for the subjects themselves (such as appraisal or reputation gains) and thus, a selfish reward. In contrast, our study examined processing of events that are exclusively relevant to someone else, i.e. a charity organization. Along with a limited number of previous studies, our study specifically addressed the processing of nonselfish, prosocial preferences. For such prosocial decisions, Hare *et al.* (2010) have demonstrated overlapping activations during value computations for personal and charity's money in the medial ventral prefrontal cortex at the decision stage. In our study, we also tested whether we could detect a value signal that would predict the subsequent choice but did not observe such a signal. It is likely that the higher variance in the values attributed to different charity organizations (in Hare *et al.* (2010), negatively evaluated organizations were included) contributed to the better detection of such signals. In addition, we tested for the effect of different decision types but did not find NAc activity or activity in the other ROIs specifically linked to donation decisions. Thus, our results do not offer support for a specific rewarding effect of the act of donating itself in the sense of a 'warm glow effect' as described previously (Moll *et al.*, 2006; Harbaugh *et al.*, 2007). This discrepancy might partly be due to the fact that we compared different types of active decisions, whereas in Harbaugh *et al.* (2007) a condition with active decision making (voluntary transfer) was compared with passive observations of transfers initiated by a computer. Our results do show activation above baseline level during all kinds of decisions. This finding is principally compatible with results of Moll *et al.* (2006), who describe common striatal activation during both pure reward decisions and NCD decisions. However, in our study activity during donation decisions was not higher than during pure reward or nondonation decisions, which leaves open the possibility that NAc activity reflects processes related to decisions in general. The comparison of different decision types in our study might be aggravated by response time differences between different decision types. In our study, donators took longer to decide in costly donation situations. In contrast, nondonators took longest in NCD situations. Such reaction time differences post a considerable problem for imaging analyses, because differences in BOLD signal might simply reflect RT differences. In our general linear model (GLM), we have included RT as duration of the respective events [which is in line with recommendations derived from empirical work by Grinband *et al.* (2008)] but nevertheless, the comparison of decisions with different RT might be problematic. Furthermore, the number of costly donation decisions varied substantially within the donator group, and in several subjects we observed less than 10 occurrences of these events, which limits the power of contrasts between costly donation decisions and other decisions. The lack of specific activity related to donation decisions might additionally be explained by the absence of social approval due to the anonymization of decisions since social approval enhances NAc activity during donation decisions (Izuma *et al.*, 2010). Finally, as a null finding, our results do not principally contradict the assumption of action orientation in charitable donations.

At the stage of outcome processing, previous studies have demonstrated increased activation in the ventral striatum for *socially* preferred outcomes, e.g. in the context of ultimatum bargaining (de Quervain *et al.*, 2004; Tabibnia *et al.*, 2008), cooperation tasks (Phan *et al.*, 2010) or inequity treatments (Fliessbach *et al.*, 2007; Tricomi *et al.*, 2010). These activations are observed even if events are neutral or negative with respect to personal monetary belongings. In the context of

charitable giving, Harbaugh *et al.* (2007) have demonstrated ventral striatal activation during passive observations of money transfers to a charity. Beyond the mere existence of such outcome-related activation, our results show that NAc activity in response to payoffs to charity is parametrically modulated by an RPE term in the same way that BOLD signals respond to one's individual rewards. Remarkably, our paradigm allowed us to observe this modulation in the same subjects and at the same time as the corresponding signal for their own monetary outcomes. Our results further demonstrate that the amount of modulation of NAc activity by the charity RPE differed inter-individually, depending on a subject's donation behavior, which in turn was linked to everyday prosocial activities. This finding is in line with results showing that NAc activity generally reflects subjective rather than objective value of rewards (Tobler *et al.*, 2007) and with more specific findings linking social value orientations and reward-related brain activity (Haruno and Frith, 2010).

For our main analysis, we averaged activity across all voxels from an anatomically defined region of interest (NAc). This was done under the assumption that the NAc is functionally homogenous. Recent data suggest a functional specialization between the NAc and the adjacent septal region, with the latter being more strongly related to social aspects of rewarding events (Moll *et al.*, 2006; Hsu *et al.*, 2008) and social attachment (Krueger *et al.*, 2007). Although the relatively low spatial resolution of fMRI must be considered, it is interesting that the peak voxels and the majority of voxels showing activations for the charity's payoff RPE were located slightly medial to the predefined NAc ROI within the septal region, whereas the peak for the personal rewards is located more laterally. The activation cluster for the charity RPE extends anterior into the subgenual area (BA 25). This finding nicely complements a previous finding by Moll *et al.* (2006), who found that decision-related reward activity in charitable giving was specifically associated with activity in the septal/subgenual region. Conversely, we found a personal payoff RPE signal in the ventral midbrain and more anterior parts of the mOFC, and here no equivalent signal for the charity's payoff was observed. Together with previous findings, our results thus suggest commonalities in the processing of not only personal and social rewards (overlapping RPE signal in the NAc), but also specific reward signals in the context of prosocial behavior, with involvement of the septal area and the subgenual part of the cingulate cortex. The specific contributions of these brain areas to social cognition are a promising target for future research.

In decision experiments, it is notoriously difficult to disentangle the effects of action and outcomes because they are typically observed simultaneously. We propose a simple but innovative manipulation to selectively test outcome-related effects: by discarding part of the subject's decisions, we introduced an RPE for given outcomes. We propose such a procedure as a generally useful method for testing outcome values in decision tasks. This approach makes explicit use of reverse inference: the observed brain signal is used to infer underlying psychological constructs such as preferences. Although many studies suggest that reward-associated signals in the NAc can serve as a surrogate marker for subjects' preferences (Knutson *et al.*, 2008), reverse inferential conclusions always need to be drawn with caution because, obviously, brain signals observed by fMRI are never unambiguous in their meaning [for a comprehensive discussion, see Poldrack (2006)]. In our case, the observed RPE signal for the charity's payoff thus suggests the effectiveness of outcome-related motives but it cannot be regarded as direct proof of such motives.

Our study design included two different events of RPE induction. One took place immediately after each decision (Session 1), the other took place after all decisions had been made (Session 2). Only the outcomes from the preceding decision experiment were shown, and they were either discarded or confirmed. We expected to detect RPEs at both time points. The second RPE induction was implemented to test whether RPE signals during Session 1 might be influenced by the previous decision. The results differ between these two RPE events. Unexpectedly, during the first RPE induction, subjects in the donator group had a higher RPE signal for their own payoff than nondonators, and only a marginally significant RPE effect for the charity payoff. The results for the second RPE event appear much clearer, with a significant effect for the charity payoff in donators and no such effect in nondonators, with a significant group difference. On the other hand, there is a similar, highly significant effect for the personal payoff RPE in both groups. We can only speculate about the reasons for these differences between the two time points. Generally, we assume that during the first part of the experiment, subjects might spend less attention to the outcomes than during the second part, where these outcomes are all they are presented with. This does not fully explain why the different RPE signals seem to be differentially affected in the two groups, i.e. why the charity RPE is lower in the donator group (compared to the second RPE induction) and the personal payoff RPE is lower in the nondonator group. Interestingly, the charity's payoff RPE observed during the decision part was higher after exclusion of costly donation condition trials. As mentioned before, the time-point of the second RPE induction appears to be the clearest test of outcome-related activity. For this event (and for the average of the two RPE inductions), there is a highly significant effect for the charity RPE in the donator group, which constitutes our main finding.

In conclusion, our results provide a first demonstration of an RPE signal in the NAc for a monetary outcome that is exclusively relevant to unrelated others. This provides additional evidence for the assumption that common brain mechanisms underlie the processing of nonselfish, social

and nonsocial rewards. The pattern of activation furthermore suggests an involvement of the septal region and the subgenual area in the processing of such rewards. Our results suggest that money belonging to a charity organization carries a reward value for subjects who are willing to donate and thereby provide neurophysiological support for the assumption of outcome-oriented motives in charitable giving.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online

Conflict of Interest

None declared.

REFERENCES

Adolphs, R. (2003). Investigating the cognitive neuroscience of social behavior. *Neuropsychologia*, *41*, 119–26.

Andreoni, J. (1989). Giving with impure altruism—applications to charity and Ricardian equivalence. *Journal of Political Economy*, *97*, 1447–58.

Andreoni, J. (1990). impure altruism and donations to public-goods—a theory of warm-glow giving. *Economic Journal*, *100*, 464–77.

Breiter, H.C., Aharon, I., Kahneman, D., Dale, A., Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron*, *30*, 619–39.

Camerer, C.F., Fehr, E. (2006). When does "economic man" dominate social behavior? *Science*, *311*, 47–52.

de Quervain, D.J., Fischbacher, U., Treyer, V., et al. (2004). The neural basis of altruistic punishment. *Science*, *305*, 1254–8.

Fliessbach, K., Weber, B., Trautner, P., et al. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science*, *318*, 1305–8.

Grinband, J., Wager, T.D., Lindquist, M., Ferrera, V.P., Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *Neuroimage*, *43*, 509–20.

Harbaugh, W.T. (1998). What do donations buy? A model of philanthropy based on prestige and warm glow. *Journal of Public Economics*, *67*, 269–84.

Harbaugh, W.T., Mayr, U., Burghart, D.R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, *316*, 1622–5.

Hare, T.A., Camerer, C.F., Knoepfle, D.T., Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience*, *30*, 583–90.

Haruno, M., Frith, C.D. (2010). Activity in the amygdala elicited by unfair divisions predicts social value orientation. *Nature Neuroscience*, *13*, 160–1.

Hsu, M., Anen, C., Quartz, S.R. (2008). The right and the good: distributive justice and neural encoding of equity and efficiency. *Science*, *320*, 1092–5.

Izuma, K., Saito, D.N., Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron*, *58*, 284–94.

Izuma, K., Saito, D.N., Sadato, N. (2010). Processing of the incentive for social approval in the ventral striatum during charitable donation. *Journal of Cognitive Neuroscience*, *22*, 621–31.

Knutson, B., Delgado, M.R., Phillips, P.E.W. (2008). Representation of subjective value in the striatum. In: Glimcher, P.W., Camerer, C.F., Fehr, E., Poldrack, R.A., editors. *Neuroeconomics*. London, San Diego, Burlington: Academic Press.

Knutson, B., Westdorp, A., Kaiser, E., Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage*, *12*, 20–7.

Konow, J. (2010). Mixed feelings: theories of and evidence on giving. *Journal of Public Economics*, *94*, 279–97.

Krueger, F., McCabe, K., Moll, J., et al. (2007). Neural correlates of trust. *Proceedings of the National Acadamy of Sciences United States of America*, *104*, 20084–9.

Lin, A., Adolphs, R., Rangel, A. (2011). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience*, doi:10.1093/scan/nsr006.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Acadamy of Sciences United States of America*, *103*, 15623–8.

Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*, 653–60.

Pagnoni, G., Zink, C.F., Montague, P.R., Berns, G.S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience*, *5*, 97–8.

Phan, K.L., Sripada, C.S., Angstadt, M., McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proceedings of the National Acadamy of Sciences United States of America*, *107*, 13099–104.

Poldrack, R.A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*, 59–63.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*, 1–27.

Smith, D.V., Hayden, B.Y., Truong, T.K., Song, A.W., Platt, M.L., Huettel, S.A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *Journal of Neuroscience*, *30*, 2490–5.

Tabibnia, G., Satpute, A.B., Lieberman, M.D. (2008). The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, *19*, 339–47.

Tobler, P.N., Fletcher, P.C., Bullmore, E.T., Schultz, W. (2007). Learning-related human brain activations reflecting individual finances. *Neuron*, *54*, 167–75.

Tricomi, E., Rangel, A., Camerer, C.F., O'Doherty, J.P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, *463*, 1089–1091.

Yacubian, J., Glascher, J., Schroeder, K., Sommer, T., Braus, D.F., Buchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *Journal of Neuroscience*, *26*, 9530–7.

Zink, C.F., Tong, Y., Chen, Q., Bassett, D.S., Stein, J.L., Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. *Neuron*, *58*, 273–83.

doi:10.1093/scan/nss147 SCAN (2014) 9, 342–349

Effort increases sensitivity to reward and loss magnitude in the human brain

Julen Hernandez Lallement,[1,2] Katarina Kuss,[1,3] Peter Trautner,[1,4] Bernd Weber,[1,3,4] Armin Falk,[1] and Klaus Fliessbach[1,3,5,6]

[1]Center for Economics and Neuroscience, University of Bonn, Nachtigallenweg, 53113 Bonn, Germany, [2]Department of Comparative Psychology, Institute of Experimental Psychology, Heinrich-Heine University Düsseldorf, 40225 Düsseldorf, Germany, [3]Department of Epileptology, University Hospital Bonn, Sigmund-Freud-Str. 25, 53105 Bonn, Germany, [4]Life & Brain Center, Department of NeuroCognition, University of Bonn, Sigmund-Freud-Str. 25, 53105 Bonn, Germany, [5]Department of Psychiatry, University Hospital Bonn, Sigmund-Freud-Str. 25, 53105 Bonn, Germany, and [6]German Center for Neurodegenerative Diseases (DZNE), Sigmund-Freud-Str. 25, 53105 Bonn, Germany

It is ecologically adaptive that the amount of effort invested to achieve a reward increases the relevance of the resulting outcome. Here, we investigated the effect of effort on activity in reward and loss processing brain areas by using functional magnetic resonance imaging. In total, 28 subjects were endowed with monetary rewards of randomly varying magnitude after performing arithmetic calculations that were either difficult (high effort), easy (low effort) or already solved (no effort). Subsequently, a forced donation took place, where a varying part of the endowment was transferred to a charity organization, causing a loss for the subject. Results show that reward magnitude positively modulates activity in reward-processing brain areas (subgenual anterior cingulate cortex and nucleus accumbens) only in the high effort condition. Furthermore, anterior insular activity was positively modulated by loss magnitude only after high effort. The results strongly suggest an increasing relevance of outcomes with increasing previous effort.

Keywords: effort; reward processing; functional magnetic resonance imaging; locus of control

INTRODUCTION

Reward processing is essential for adaptive behavior. Rewards initiate approach behavior, induce reward-related learning, guide decision and lead to positive hedonic feelings (Schultz, 2006). In natural environments, rewards rarely occur without previous effort, such as effort invested in foraging activities. Ecological theories hold that effort is incorporated into the processing of the resulting outcome to guide future decisions (Stephens and Anderson, 2001; Kolling *et al.*, 2012). In humans (Prévost *et al.*, 2010) and in animals (Rudebeck *et al.*, 2006; Walton *et al.*, 2007), recent findings indicate that effortful options are devaluated, suggesting that effort is incorporated as a cost when a decision is faced (effort discounting during valuation of choice alternatives). Only a few studies, however, have addressed the effect of prior effort on the processing of a subsequent reward. Animal studies have demonstrated that expenditure of effort increases the hedonic value of the resulting reward in a variety of species (Lewis, 1964; Clement *et al.*, 2000; Johnson and Gallagher, 2011). In humans, prior actions or skill demand can increase the valuation of a subsequent reward (Zink *et al.*, 2004; Alessandri *et al.*, 2008; Vostroknutov *et al.*, 2012). Further insights on the effect of effort on outcome valuation come from behavioral economic studies: The willingness to spend money is decreased for earned gains in comparison with non-earned ('windfall') gains (Muehlbacher and Kirchler, 2009). In contrast to these studies showing an increase in value for rewards following effort, Botvinick *et al.* (2009) report a neural activity pattern consistent with the concept of effort discounting; i.e. a devaluation of rewards following effort. In this article, we investigate how prior effort influences the valuation of different reward magnitudes. We do so under the assumption that effort does not constantly increase or decrease the valuation of a subsequent reward, but that it has differential effects on different levels of rewards: relatively low rewards might become devaluated, whereas the value of relatively high rewards might increase. In other words, the sensitivity to reward magnitude might increase with increasing effort. Such a mechanism appears plausible for organisms to assess whether an effort 'paid off'. As we assumed that this applies to outcomes in general we also investigated the effect of effort on subsequent losses of varying magnitude.

Because the neurophysiological correlates of outcome processing are well characterized, we investigated the effect of effort on reward processing using functional magnetic resonance imaging (fMRI). While the subjects were being scanned, we provided them with monetary rewards and varied the amount of effort they had to invest to solve the prior task. Subsequently, we induced a forced donation event, in which the endowment was randomly divided between the subject and a charity organization, causing a personal loss for the subject. Our study is able to provide new insights into the effect of effort on the neural processing of outcomes, because both reward and loss magnitudes were varied over a large range.

Reward processing is a major function of the mesolimbic dopaminergic system, which includes the projection sites of dopaminergic midbrain neurons to the nucleus accumbens (NAcc) (Breiter *et al.*, 2001) and to the medial prefrontal cortex including the subgenual anterior cingulate cortex (sgACC) and the anterior medial orbitofrontal cortex (mOFC) (O'Doherty *et al.*, 2003). Neuroimaging studies suggest a functional specialization in the components of this network, with the NAcc rather encoding reward prediction errors (RPEs, i.e. the difference between actual and expected outcome) and the medial prefrontal cortex rather encoding absolute reward value (Rangel and Hare, 2010). This functional specialization is of minor concern in this study because our paradigm neither includes reinforcement learning (which would demand RPE representations) nor decision making (demanding value representations). As both RPE and reward value increase with increasing reward magnitude we expected a positive correlation between

Received 26 July 2012; Accepted 26 November 2012
Advance Access publication 30 November 2012
The first two authors contributed equally to this work.
J. H. L., K.K. and K.F. are funded by the German Research Council (Grant FL 715/1-1). B.W. is funded by the German Research Council with a Heisenberg Grant (Grant WE 4427/3-1).

Correspondence should be addressed to Klaus Fliessbach, MD, Department of Psychiatry, University of Bonn Medical Center, Sigmund Freud-Str. 25, D-53127 Bonn, Germany. E-mail: klaus.fliessbach@ukb.uni-bonn.de

reward magnitude and the blood oxygen level-dependent (BOLD) signal in both areas. Our research question was whether BOLD responses to the same rewards depend on the history of their acquisition (e.g. effortful or not). The neural correlates of loss processing are less well defined, but some key areas have consistently been implicated in loss processing: the anterior insula (AI), the anterior cingulate cortex and the amygdala (see Liu *et al.*, 2011 for a meta-analysis).

We hypothesized that neural activity in reward (loss) processing areas should depend more strongly on reward (loss) magnitude in the high effort condition. Importantly, this does not necessarily imply that a reward as such is valued more after supplying effort than after providing no effort, but that large rewards are valued more in comparison to low rewards especially after high effort (i.e. effortful gains have a steeper utility function). Accordingly, we expected that increased effort leads to: (i) stronger effects of reward magnitude on activity in the NAcc and basal medial prefrontal areas at reward receipt and (ii) stronger effects of the loss magnitude on activity during the forced donation event in loss-related structures, such as AI, dorsal anterior cingulate cortex (dACC) and amygdala.

MATERIALS AND METHODS

Participants and procedure

Thirty subjects (14 female, mean age = 25.4 years, s.d. = 4.03 years) participated in the fMRI experiment, two participants were excluded from analysis due to excessive head movement. All subjects were native German speakers, right-handed and had no history of psychiatric or neurological disorders. During the experiment, subjects were presented with a total of 120 arithmetical tasks that were either difficult (40, high effort condition), easy (40, low effort condition) or already solved (40, no effort condition). After 8 s time for the calculation, subjects were presented with four possible solutions and had to quickly (3 s) choose the correct answer by button press (Figure 1). The limited time for the answer aimed at forcing subjects to actually perform the calculation when the task was displayed.

In case of an incorrect response, a negative visual feedback (red 'X') was displayed, and the next trial started. After a correct response, the subjects were first presented with a positive visual feedback (green check mark). Displaying a correctness informing feedback at this point is crucial to avoid uncertainty about reward reception for the following endowment event. After this visual feedback, subjects were endowed with money ('endowment' event). Subsequently, the endowment was split between the subject and a charity organization ('forced donation' event).

Subjects had no control over the monetary donations; instead they were presented with a forced donation. All endowments and splits were randomly determined by the experiment software. The endowments ranged from 5€ to 35€, in steps of 1€. Subjects were informed about the maximum endowment before the experiment. The amount taken away during the forced donation event ranged from 0€ to the entire endowment amount. Crucially, randomization of endowments and splits was across the three effort conditions, so that the average reward and loss magnitude was randomly distributed between conditions. At the end of the experiment, one trial was randomly selected and implemented: charity and participant (additionally to a 15€ show up fee) received the payoffs of this trial, but only if this trial was solved correctly. Subjects received detailed written and verbal instructions and gave informed written consent. The study was approved by the Ethics committee of the University of Bonn.

Additionally, 30 different subjects performed a behavioral study (see Supplementary Methods). Contrary to the fMRI paradigm, these subjects were offered the possibility to donate money obtained through the same conditions presented as in the fMRI paradigm. This manipulation allowed us to test whether our experimental manipulation would actually influence monetary decisions.

Locus of control questionnaire

After the fMRI experiment, subjects completed a German version of the locus of control questionnaire (Krampen, 1981). The questionnaire consists of three subscales which asses beliefs related to self-efficiency

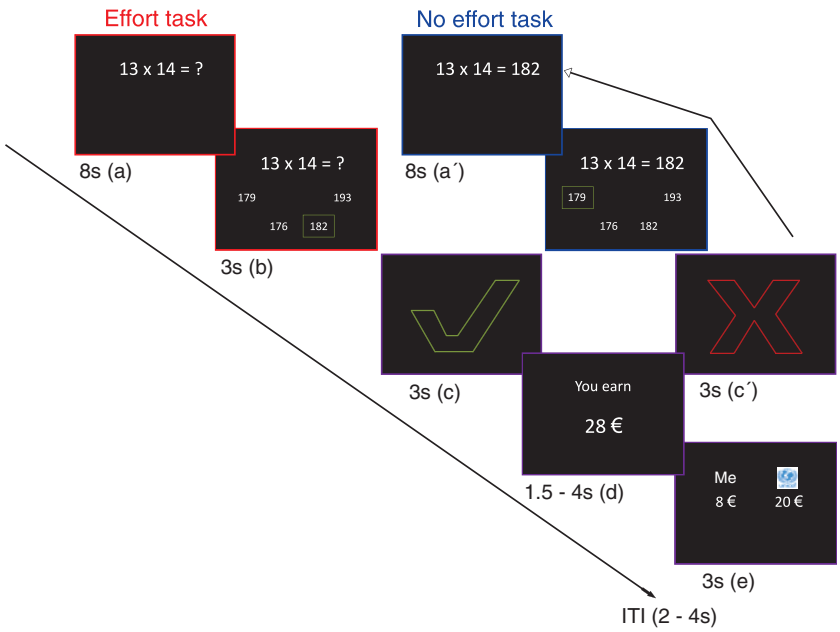

Fig. 1 Timeline of the experiment. Subjects were confronted with a calculation task, which they either solved on their own (effort task) or which was already presented with the correct solution (no effort task) (a). Subjects chose one solution out of four options. The chosen option was highlighted with a rectangle (b). A feedback signaled the correctness of the answer (c). Only in case of correct solution, subjects received an endowment ('endowment'event) (d); otherwise (c') a new trial started. A random split of the endowment (between subject and charity) was presented ('forced donation' event) (e). A jittered inter-trial interval (2–4 s) followed. Note that the sequence was identical for effort and no effort, except for the already displayed solution in the no-effort condition.

and causal attribution (for details on the subscales, see Methodological Details in Supplementary Material). We implemented this questionnaire because our experimental design aims at inducing an internal attribution of a reward in the high effort condition. Therefore, the effects of effort on neural reward- and loss-processing should depend on the predominant attribution style of the subjects.

fMRI scanning and analysis

Scanning was performed on a 1.5-T Avanto Scanner (Siemens, Erlangen, Germany) using an 8-channel head coil. Functional data were acquired using Echo Planar Imaging (EPI)-sequences (for more information, see Methodological Details in Supplementary Material).

Brain imaging analyses focused on changes in activation during endowment presentation (Figure 1d) and during forced donation (Figure 1e). For both events, we included parametric modulators in the first level general linear model (GLM), that represent the total reward magnitude (endowment event) and the relative loss magnitude (forced donation) of a given trial, respectively. The relative loss was defined as the ratio of the charity's payoff to the reward magnitude of the endowment (i.e. the fraction of money taken away from the subject). See Methodological Details in Supplementary Material for a detailed description of the GLM.

Single-subject contrasts were computed for each regressor. These were subjected to a random effects second-level analysis to test main effects of the parameters 'reward magnitude' and 'loss magnitude' for each single condition (one sample *t*-tests), to test differences between conditions (within subjects ANOVAs) and to test for correlations with questionnaire data (see Methodological Details in Supplementary Material).

For all group level analyses, we applied an inclusion threshold of $P = 0.005$ uncorrected, and performed Family wise error (FWE) correction for multiple comparisons restricted to the regions of interest (ROI) (endowment event) or for the whole brain (forced donation event). To test for overlapping effects of the contrasts high effort > no effort and high effort > low effort, we performed an inclusive masking procedure as implemented in SPM8. We tested the effect of high effort > no effort on $P = 0.005$ (uncorrected) and masked this with the effect of high effort > low effort [mask image $P = 0.05$ (uncorrected)]. The same masking procedure was conducted, using a mask derived from the contrast low effort > no effort. This masking procedure resembles a conjunction analysis with a logical 'AND'. In contrast to a standard conjunction analysis as implemented in SPM8, it uses different thresholds for different contrasts. Note that the resulting clusters of activation were tested on a FWE-corrected *P*-value of 0.05, either small volume corrected for the ROI or on the whole-brain level.

Regions of interest definition

We defined areas that are known to be involved in reward processing (NAcc, sgACC and mOFC) as ROI for the analysis of the endowment event. As the processing of loss events is less clearly restricted to singular regions, we refrained from defining a priori ROI and analyzed this data on the whole-brain level. After we identified a whole-brain correctable effect for the high effort condition in the frontal operculum/insula, we *post hoc* included an anatomically defined mask of the insula, to focus on differential effects of the conditions in this region.

RESULTS

Task effects on performance and brain activation

Accuracy was 81.36% (s.d. = 12.12%) in the high effort condition, 94.17% (s.d. = 6.2%) in the low effort condition and 98.49% (s.d. = 2.07%) in the no effort condition, yielding a significant condition effect on accuracy ($F = 55.098$, $P < 0.001$). The lowest number of correct solutions in the high effort condition was 22, yielding a sufficient number of events entering the fMRI analysis. Comparing task related activity for the effort conditions (low + high) with the no effort condition yielded activity in areas known to be related to arithmetic processing, such as the intraparietal sulci, inferior prefrontal cortices and precentral cortices (Dehaene *et al.*, 2004). In a widely overlapping network activity was higher for the high effort than for the low effort condition (see Tables S1–S3 and Figure S3).

Activity in the sgACC and NAcc scales with reward magnitude after high effort

At the time of the endowment presentation, we observed a significant positive modulation of the BOLD signal by the amount of money earned for the high effort condition in the sgACC [MNI coordinates of peak voxel: $X = 6$, $Y = 23$, $Z = -23$, $t = 4$, pFWE (small-volume corrected) < 0.05]. There was no such association in the other conditions. Based on the hypothesis that reward sensitivity is highest after high effort, we tested for differences between conditions (high effort > no effort, high effort > low effort and low effort > no effort). We observed a significant stronger modulation of BOLD signal by the reward amount after high effort compared with no effort in the sgACC [$X = 3$, $Y = 20$, $Z = -23$, $t = 4.53$, pFWE (small-volume corrected) < 0.05]. There was no such effect for the contrast high effort > low effort or low effort > no effort, when we applied the same threshold of $P = 0.005$, uncorrected. To test whether differences between the other conditions exist on a more lenient threshold and to test for overlapping activity for the contrasts, we performed an inclusive masking procedure (as described in the 'Materials and Methods' section). This procedure demonstrates overlapping effects for the contrast (high effort > no effort and high effort > low effort) in the sgACC surviving small volume correction [$X = 3$, $Y = 20$, $Z = -23$, $t = 4.53$, pFWE (small-volume corrected) < 0.05], see Figure 2. For the contrast low effort > no effort, no overlapping effect emerged. Figure 3a shows the mean parameter estimates for the parametric modulator 'reward magnitude' in the three experimental conditions averaged across a 5-mm sphere in the sgACC ($X = 3$, $Y = 20$, $Z = -23$). This figure serves demonstration reasons only, no statistical inferences are based on these averaged parameters.

We observed the same pattern in the NAcc: positive modulation of BOLD signal with the amount of money earned only in the high effort condition [MNI coordinates of peak voxel: $X = 9$, $Y = 11$, $Z = -5$, $t = 3.5$, pFWE (small-volume corrected) < 0.05], see also Figure 3b. Single voxels in the right NAcc showed significant differences surviving small volume correction for the following contrasts: high effort > no effort [$X = 6$, $Y = 11$, $Z = -5$, $t = 2.8$, pFWE (small-volume corrected) < 0.05] and high effort > low effort [$X = 6$, $Y = 8$, $Z = -5$, $t = 3.07$, pFWE (small-volume corrected) < 0.05]. Applying our inclusive masking procedure, we identified overlapping effects of both contrasts surviving small volume correction in the right NAcc [$X = 6$, $Y = 11$, $Z = -5$, $t = 2.8$, pFWE (small-volume corrected) < 0.05]. Again, no overlapping effects of the contrast low effort > no effort was observed. There was no significant effort related effect in the mOFC although on a more lenient threshold ($P < 0.005$, uncorrected) the same pattern emerged. In sum, these findings demonstrate a stronger association of BOLD signal with reward magnitude in reward-processing areas (sgACC and NAcc) specifically in the high effort condition. These results do not imply that the BOLD signal is in absolute terms higher in the high effort condition. To illustrate that point, we ran additional analysis splitting up the endowment event in four onset regressors according to the reward amount (very low, low, medium and high) for each experimental condition (see Additional Results and Figure S6 in Supplementary Material), showing a linear

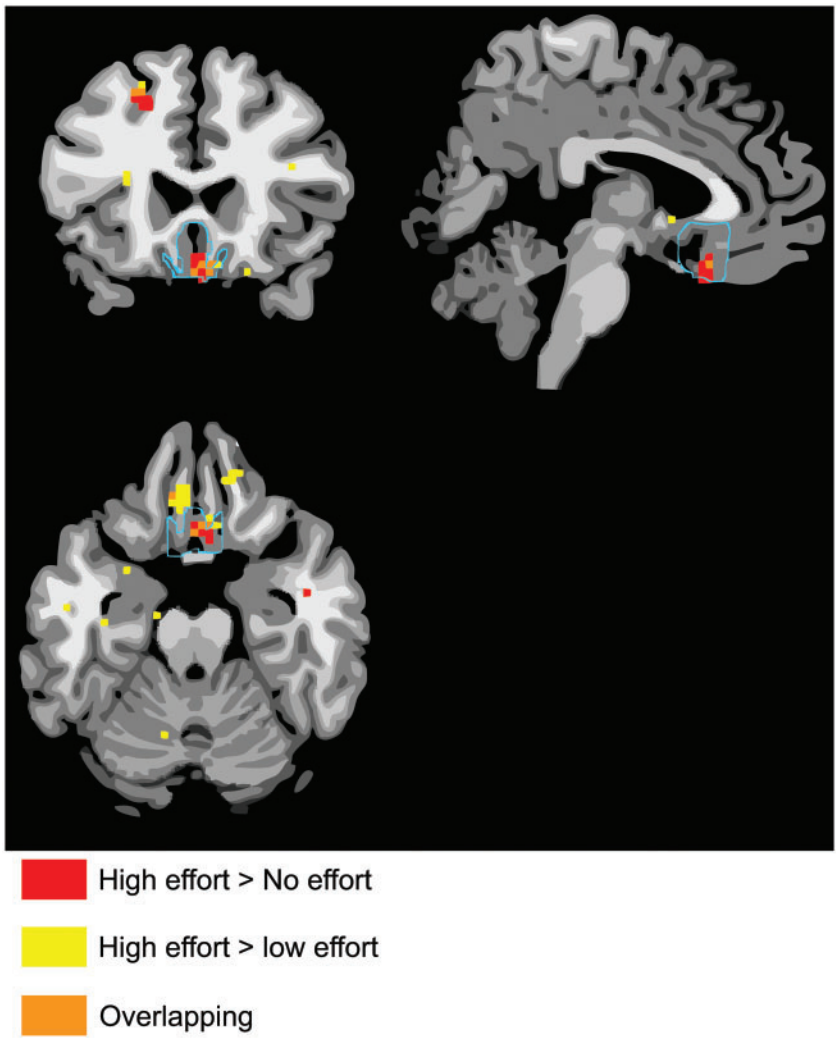

Fig. 2 Subgenual ACC. Overlapping effects (orange) of the contrasts high effort > no effort (red) and high effort > low effort (yellow). The sgACC is framed in blue. There was a stronger positive modulation of the BOLD signal by the endowment after high effort (high effort > no effort in red, thresholded at $t > 2.67$, corresponding to $P < 0.005$, uncorrected; high effort > low effort in yellow, thresholded at $t > 2.39$, corresponding to $P < 0.01$, uncorrected). MNI coordinates: $X = 3$, $Y = 23$, $Z = -21$.

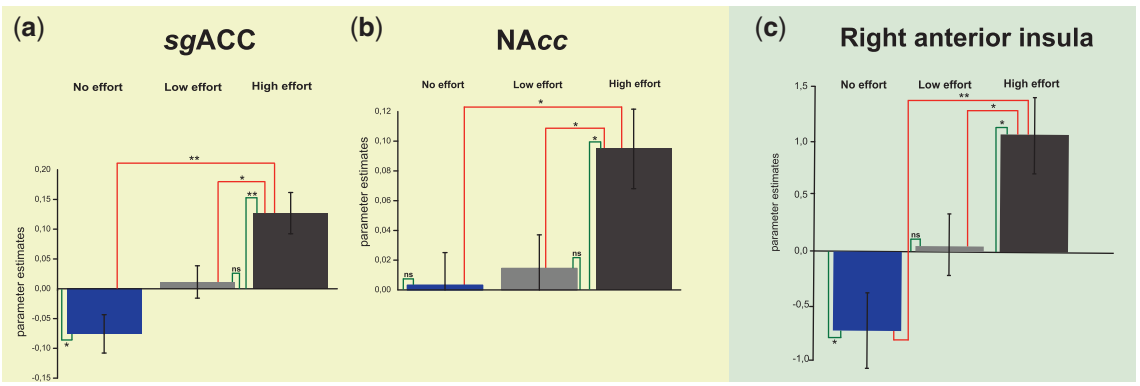

Fig. 3 Reward-related activity in the subgenual ACC (a) and NAcc (b), as well as loss-related activity in the AI (c) for each condition. (**a**) Mean parameter estimates (±s.e.m.) for the parametric modulator of the endowment averaged across a 5-mm sphere in the subgenual ACC ($X = 3$, $Y = 20$, $Z = -23 + 5$ mm). (**b**) Mean parameter estimates (±s.e.m.) for the parametric modulator of the endowment averaged across a 5-mm sphere in the NAcc ($X = 6$, $Y = 11$, $Z = -5 + 5$ mm). (**c**) Mean parameter estimates (±s.e.m.) for the parametric modulator of the relative loss averaged across 5-mm sphere in the AI/frontal operculum ($X = 39$, $Y = 8$, $Z = -8 + 5$ mm). The respective voxels are the peakvoxels of the overlapping effect of the contrast high effort > no effort AND high effort > low effort (identified by masking procedure). These barplots serve demonstration reasons only, no statistical inferences are based on those barplots. $^{**}P < 0.001$; $^{*}P < 0.05$; P's are two-tailed.

relation of the BOLD signal with the reward/loss magnitude only in the high effort condition.

Activity in the insula scales with relative loss after high effort

At the presentation of the forced donation, we observed a significant positive modulation of the BOLD signal by the relative amount of money taken away from the subject (relative loss) in the AI/frontal operculum only for the high effort condition [Left peak voxel: $X=-51$, $Y=11$, $Z=4$, $t=4.53$, pFWE (whole brain, cluster level) <0.05 with 106 voxels; right peak voxel: $X=42$, $Y=2$, $Z=-11$, $t=4.00$, pFWE (whole brain, cluster level) <0.05 with 111 voxels]. These two activation clusters are the only clusters surviving whole brain correction for multiple comparisons on cluster-level [P (FWE) <0.05]. In the other conditions, no such association was found. We tested an alternative GLM with the absolute loss as the parametric modulator of the split event for each condition (instead of the relative loss). There is no significant modulation of BOLD signal by the absolute loss amount in neither condition (high, low and no effort).

Figure 3c shows the mean parameter estimates for the parametric modulator (relative loss) in the three experimental conditions averaged across a 5-mm sphere in the AI ($X=39$, $Y=8$, $Z=-8$). We tested for differences between conditions (high effort > no effort; high effort > low effort). We observed a significantly stronger modulation of the BOLD signal by the relative loss after high effort than after low effort in the AI/frontal operculum [MNI coordinates: $X=-51$, $Y=11$, $Z=4$, $t=5.07$, pFWE (whole brain/cluster level) <0.05, 189 voxels in the cluster]. The contrast high effort > no effort yields similar activation clusters in the bilateral frontal operculum and insula, surviving small volume correction for multiple comparisons in the insula [MNI coordinates: $X=39$, $Y=8$, $Z=-8$, $t=4.47$, pFWE (small-volume corrected) <0.05]. Figure 4 shows overlapping activity for both contrasts (high effort > low effort; high effort > no effort). We applied the same inclusive masking procedure as for the payoff event to test for overlapping activity for the contrasts (high effort > no effort and high effort > low effort). This procedure demonstrates overlapping effects in an anatomically defined insula mask surviving small volume correction [$X=39$, $Y=8$, $Z=-8$, $t=4.47$, pFWE (small-volume corrected) <0.05]. No such effects emerged for the contrast low effort > no effort. In sum, these findings indicate that an increase in neural activity with the magnitude of monetary losses in the insula is

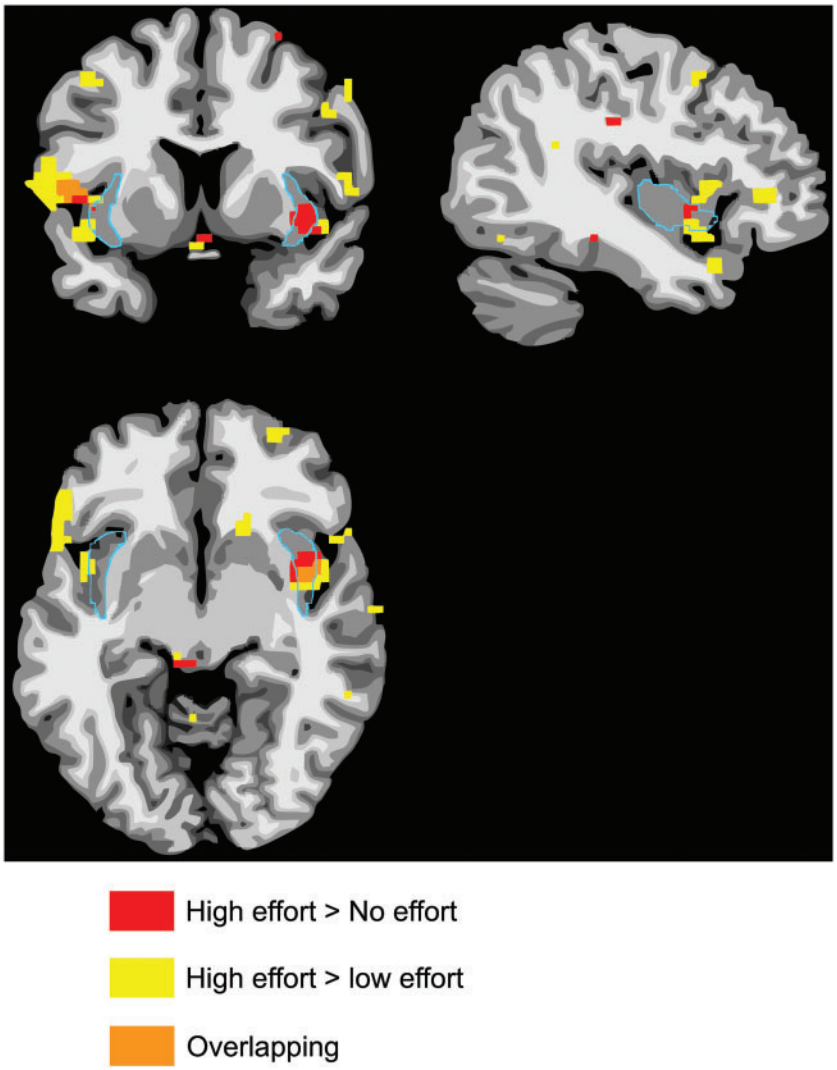

Fig. 4 AI. Overlapping effects (orange) of the contrasts high effort > no effort (red) and high effort > low effort (yellow). The insula is framed in blue. There was a stronger positive modulation of the BOLD signal by the relative loss after high effort (high effort > no effort in red, thresholded at $t>2.67$, corresponding to $P<0.005$, uncorrected; high effort > low effort in yellow, thresholded at $t>2.49$, corresponding to $P<0.008$, uncorrected). MNI coordinates: $X=-42$, $Y=9$, $Z=-8$.

specific for the high effort condition. Neither the amygdala nor the dACC showed a significant loss-magnitude related activation.

To further illustrate the differences between the high effort and the other two conditions, we performed an additional analysis for the contrast high effort > low + no effort, showing significant (FWE corrected) differences between these conditions in the sgACC and NAcc for the endowment event and bilaterally in the insula for the loss event (see Figures S4 and S5).

Behavioral study: effort changes active donation behavior

Consistent with the neural pattern of increased insular sensitivity to relative loss magnitude we observed that our experimental conditions influence active donation decisions in our additional behavioral experiment (see Additional Results and Figures S1 and S2 in Supplementary Material). Donation rates were lowest in the high effort condition. Comparing means between the high effort and the no effort condition revealed a strong trend for lower donations after high effort (dependent samples *t*-test, $P = 0.06$, one-sided). The number of subjects that donated more after high effort than after no effort was significantly lower than the number of subjects showing the opposite pattern (sign-test, $P = 0.022$, one-sided); see Figure S2.

The effect of effort on insular loss processing is modulated by personal attribution style

We extracted the parameter estimates (parametric modulators for reward and relative loss) for the peak voxels (+5 mm sphere) showing overlapping effects for high effort > no effort and high effort > low effort, and correlated the parameters with the subscales of the locus of control questionnaire and with a difference score between the scales internality and chance. This yields 16 comparisons (4 locations × 4 scales), demanding a significance level of $P < 0.003$, to test on an overall error probability of 0.05. On this significance level, there was a positive correlation ($r = 0.574$, $P = 0.001$, two-sided) of the loss-related activity in the left insula/frontal operculum and the scale chance (scale of the locus of control questionnaire, which measures the tendency of a person to attribute causes to chance or luck). Furthermore, the difference between the two scales internality and chance significantly correlated with the loss-related activity ($r = -0.593$, $P = 0.001$, two-sided).

DISCUSSION

Our study demonstrates effects of reward magnitude on reward-related brain activity in the sgACC and in the NAcc only for money gained with high effort but not for low- or non-effort gains. Conversely, we observed that a loss magnitude related signal in the AI was significantly increased after high-effort gains compared with low- or non-effort gains. These results show that reward and loss processing in the human brain critically depend on the history of a reward.

Both the sgACC and NAcc have been implicated in the processing of rewards. Rather than any measure of objective reward value, it has been shown that activity in these brain regions reflects subjective reward value, which depends on several contextual factors, e.g. previous (Elliott *et al.*, 2000) or alternative outcomes (Breiter *et al.*, 2001), the reward of others (Fliessbach, 2007) and personal assets (Tobler *et al.*, 2007). Only a limited number of studies have addressed the effect of the action preceding a reward on the processing of that reward. Zink *et al.* (2004) showed that an active choice before a reward is followed by a higher subsequent NAcc signal than for passively receiving a reward in a probabilistic reinforcement learning paradigm. This finding indicates that the *salience* of a reward, i.e. how much it is relevant for future actions, affects reward signals in the NAcc. Another recent study demonstrates higher reward signals in the medial orbitofrontal cortex after a skill demanding task compared to a luck-dependent task (Vostroknutov *et al.*, 2012). As in our study, the magnitude of the monetary reward was varied in this study. In contrast to our study, however, the payoff depended on the subject's performance. Therefore, reward activity in this condition could also result from a feedback about performance. Our study demonstrates that effort prior to a reward increases the sensitivity to monetary rewards even though reward magnitude was not related to the amount of effort.

Effort enters the decision process as a cost (Walton *et al.*, 2007), and is consistently related to decreased reward related brain activity at the decision phase (Croxson *et al.*, 2009; Kurniawan *et al.*, 2010; Prévost *et al.*, 2010). Botvinick *et al.* (2009) showed that not only at the stage of decision making but also at the stage of a receipt of a reward there was a stronger BOLD signal for low effort trials compared with high effort trials. The authors interpret this as an effort discounting pattern in the NAcc. However, this study did not parametrically vary reward magnitude and thus can only address absolute levels of reward-related BOLD signals for a constant reward magnitude. Rather than such absolute signals, our results show that the BOLD signal is more strongly related to reward magnitude after effort. It is important to note that this does not automatically imply that monetary gains after effort generally induce higher signals as illustrated by additional analyses (see Additional Results and Figure S6 in Supplementary Material). This analysis reveals that reward signals are not generally lower after effortless gains. Rather, the modulation of the BOLD signal by reward magnitude is increased after high effort. Therefore, our results do not contradict studies showing decreased reward-related brain activity in effort-related decision tasks. We rather suggest that subjects become more sensitive to differences in reward magnitude after effort. One possible (but speculative) explanation for our result is that effort induces expectancies about the following outcome. The actually occurring outcomes are then related to this expectancy, i.e. they become reference dependent. After no or low effort, such expectancies might be missing or weaker and therefore, subjects might become relatively indifferent against outcome magnitude. Therefore, our findings are consistent with the idea of reference dependency of both utility (Tversky and Kahneman, 1991) and reward-related brain activity (Fliessbach *et al.*, 2007; Tobler *et al.*, 2007). Given that we did not find significant associations between BOLD signals and reward magnitude in the other conditions one might even speculate that a clear reference point for the evaluation of the resulting outcome only emerges if previous effort was undertaken.

This interpretation is further supported by the finding of a loss magnitude effect in the AI that was specific for the high effort condition. A wealth of empirical findings demonstrates the role of the anterior insular cortex in the processing of negative emotions (Seymour *et al.*, 2005; Pessiglione *et al.*, 2006). In human studies of decision making, insular activity has also been shown in the context of outcome evaluation (Diekhof *et al.*, 2012). At the anticipatory stage, AI activity negatively scales with expected value (Rolls *et al.*, 2008) and positively with the expectation of an aversive stimulus (Seymour *et al.*, 2004). At the stage of outcome evaluation, insular activity signals both regret and disappointment (Chua *et al.*, 2009), and is predictive for behavioral changes occurring after the experience of negative outcomes (O'Doherty *et al.*, 2003). Importantly, insular BOLD signal changes scale with the magnitude of a loss (relative to an alternative outcome) in the same way as NAcc and mPFC activity reflects reward magnitude (Kuhnen and Knutson, 2005). In line with those findings, we found an association between BOLD signal magnitude and the relative loss occurring at the forced donation. No such association was found with the absolute amount of money that was lost. Based on standard utility theory, we expected that the relative loss would predominantly

determine the subjective loss in utility, as the initial endowment sets a reference-point ('how much can be lost?') to which the magnitude of the loss is related. Our results are therefore consistent with previous reports which support the view that neural activity codes subjective, reference-dependent values rather than absolute values (Rangel and Hare, 2010). Based on these findings we interpret the AI activity occurring at the stage of the forced donation as reflecting the averseness associated with the loss of the subject's personal endowment. Notably, the forced donation event also implies a gain of money for the charity. This means that the overall utility of the outcome presented at the loss event might not completely be reflected by the own personal loss. Especially in subjects with strong prosocial preferences the gain for the charity might carry its own utility, as suggested by a recent study (Kuss *et al.*, 2013). Since we did not formally test for prosocial preferences in the fMRI group, we cannot control for this factor. However, from the donation behavior in our behavioral study and from the mentioned study it appears that the majority of subjects prefer their own gains over gains for the charity. Therefore, we assumed that the utility loss occurring at the loss event is predominantly reflected by the own personal loss. Note that we did not find any NAcc significant activation related to the forced donation event. This is consistent with findings showing higher NAcc activation for voluntary acts of giving in contrast to forced donations (Harbaugh *et al.*, 2007). The finding that insular activity is modulated by the magnitude of the loss in the high effort condition but not in the two other conditions therefore complements the finding from the endowment event. We infer that the sensitivity of subjects to loss magnitude is increased for effortfully gained money.

This effect in the AI was most pronounced in subjects with high external control perception. Individuals with this personality trait tend to attribute outcomes to external causes like chance. One might speculate that these subjects are more prone to a manipulation that explicitly induces an internal attribution like the high effort condition in our study. More generally, this personality trait is related to reduced cognitive control of emotions and to an increased sensitivity for punishment (Declerck *et al.*, 2006). Therefore, the increased insular effect in externally attributing subjects could reflect a higher sensitivity of an aversive event, i.e. the loss of effortfully earned money.

Our results bear important implications. First, they suggest that the attribution of rewards to internal or external cause critically influences their evaluation as suggested by attribution theory (Wittig *et al.*, 1981; Weiner, 2000). This is further supported by the influence of the locus of control-personality trait which specifically features the predominant attribution style of persons. Second, together with previous behavioral studies, our results inform behavioral economic studies on social preferences and preferences in general about the fact, that non-earned (windfall) money is differently evaluated than earned money. Third, our results are linked to the psychological concept of deservingness, defined as the balance between outcome and action that led to it (Feather *et al.*, 2011). We speculate that the increased sensitivity to the reward and loss magnitude observed after previous effort contributes to the mechanism through which deservingness emerges. Finally, the results imply that outcomes become especially important when a preceding effort had to be taken. This assumption makes sense from an ecological point of view: if a person invests energy to obtain a reward, it should matter more whether the effort actually 'paid off'. The fact that neural correlates of this effect can be induced instantaneously by a simple experimental manipulation suggests that this is a very fundamental mechanism which might have strong biological foundations.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

REFERENCES

Alessandri, J., Darcheville, J.-C., Delevoye-Turrell, Y., Zentall, T.R. (2008). Preference for rewards that follow greater effort and greater delay. *Learning & Behavior*, *36*(4), 352–8.

Botvinick, M.M., Huffstetler, S., McGuire, J.T. (2009). Effort discounting in human nucleus accumbens. *Cognitive, Affective & Behavioral Neuroscience*, *9*(1), 16–27.

Breiter, H.C., Aharon, I., Kahneman, D., Dale, A., Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron*, *30*(2), 619–39.

Chua, H.F., Gonzalez, R., Taylor, S.F., Welsh, R.C., Liberzon, I. (2009). Decision-related loss: regret and disappointment. *NeuroImage*, *47*(4), 2031–40.

Clement, T.S., Feltus, J.R., Kaiser, D.H., Zentall, T.R. (2000). "Work ethic" in pigeons: reward value is directly related to the effort or time required to obtain the reward. *Psychonomic Bulletin & Review*, *7*(1), 100–6.

Croxson, P.L., Walton, M.E., O'Reilly, J.X., Behrens, T.E.J., Rushworth, M.F.S. (2009). Effort-based cost-benefit valuation and the human brain. *The Journal of Neuroscience*, *29*(14), 4531–41.

Declerck, C.H., Boone, C., De Brabander, B. (2006). On feeling in control: a biological theory for individual differences in control perception. *Brain and Cognition*, *62*(2), 143–76.

Dehaene, S., Molko, N., Cohen, L., Wilson, A.J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology*, *14*(2), 218–24.

Diekhof, E.K., Kaps, L., Falkai, P., Gruber, O. (2012). The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude—an activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. *Neuropsychologia*, *50*(7), 1252–66.

Elliott, R., Friston, K.J., Dolan, R.J. (2000). Dissociable neural responses in human reward systems. *The Journal of Neuroscience*, *20*(16), 6159–65.

Feather, N.T., McKee, I.R., Bekker, N. (2011). Deservingness and emotions: testing a structural model that relates discrete emotions to the perceived deservingness of positive or negative outcomes. *Motivation and Emotion*, *35*(1), 1–13.

Fliessbach, K., Weber, B., Trautner, P., et al. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. *Science*, *318*(5854), 1305–8.

Harbaugh, W.T., Mayr, U., Burghart, D.R. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science*, *316*(5831), 1622–5.

Johnson, A.W., Gallagher, M. (2011). Greater effort boosts the affective taste properties of food. *Proceedings of the Biological sciences/The Royal Society*, *278*(1711), 1450–6.

Kolling, N., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S. (2012). Neural mechanisms of foraging. *Science*, *336*(6077), 95–8.

Krampen, G. (1981). IPC-Fragebogen zu Kontrollüberzeugungen (German version of Locus of Control Questionnaire by Hanna Levenson, 1972) Göttingen: Hogrefe.

Kuhnen, C.M., Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, *47*(5), 763–70.

Kurniawan, I.T., Seymour, B., Talmi, D., Yoshida, W., Chater, N., Dolan, R.J. (2010). Choosing to make an effort: the role of striatum in signaling physical effort of a chosen action. *Journal of Neurophysiology*, *104*(1), 313–21.

Kuss, K., Falk, A., Trautner, P., Elger, C.E., Weber, B., Fliessbach, K. (2013). A reward prediction error for charitable donations reveals outcome orientation of donators. *Social Cognitive and Affective Neuroscience*, *8*(2), 216–23.

Lewis, M. (1964). Some nondecremental effects of effort. *Journal of Comparative and Physiological Psychology*, *57*(3), 367–72.

Liu, X., Hairston, J., Schrier, M., Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1219–36.

Muehlbacher, S., Kirchler, E. (2009). Origin of endowments in public good games: the impact of effort on contributions. *Journal of Neuroscience, Psychology, and Economics*, *2*(1), 59–67.

O'Doherty, J., Critchley, H., Deichmann, R., Dolan, R.J. (2003). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *The Journal of Neuroscience*, *23*(21), 7931–9.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–5.

Prévost, C., Pessiglione, M., Météreau, E., Cléry-Melin, M.-L., Dreher, J.-C. (2010). Separate valuation subsystems for delay and effort decision costs. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(42), 14080–90.

Rangel, A., Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology*, *20*(2), 262–70.

Rolls, E.T., Grabenhorst, F., Parris, B.A. (2008). Warm pleasant feelings in the brain. *NeuroImage*, *41*(4), 1504–13.

Rudebeck, P.H., Walton, M.E., Smyth, A.N., Bannerman, D.M., Rushworth, M.F.S. (2006). Separate neural pathways process different decision costs. *Nature Neuroscience*, *9*(9), 1161–8.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, *57*, 87–115.

Seymour, B., Doherty, J.P.O., Dayan, P., et al. (2004). *Temporal difference models describe higher-order learning in humans*, Vol. 429(June), 664–7.

Seymour, B., O'Doherty, J.P., Koltzenburg, M., et al. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature Neuroscience*, *8*(9), 1234–40.

Stephens, D.W., Anderson, D. (2001). The adaptive value of preference for immediacy: when shortsighted rules have farsighted consequences. *Behavioral Ecology*, *12*(3), 330–9.

Tobler, P.N., Fletcher, P.C., Bullmore, E.T., Schultz, W. (2007). Learning-related human brain activations reflecting individual finances. *Neuron*, *54*(1), 167–75.

Tversky, A., Kahneman, D. (1991). Loss aversion in riskless choice: a reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–61.

Vostroknutov, A., Tobler, P.N., Rustichini, A. (2012). Causes of social reward differences encoded in human brain. *Journal of Neurophysiology*, *107*(5), 1403–12.

Walton, M.E., Rudebeck, P.H., Bannerman, D.M., Rushworth, M.F.S. (2007). Calculating the cost of acting in frontal cortex. *Annals of the New York Academy of Sciences*, *1104*, 340–56.

Weiner, B. (2000). Intrapersonal and interpersonal theories of motivation from an attributional perspective. *Educational Psychology Review*, *12*(1), 1–14.

Wittig, M.A., Marks, G., Jones, G.A. (1981). Luck versus effort attributions: effect on reward allocations to self and other. *Personality and Social Psychology Bulletin*, *7*(1), 71–8.

Zink, C.F., Pagnoni, G., Martin-Skurski, M.E., Chappelow, J.C., Berns, G.S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron*, *42*(3), 509–17.