UNIVERSITAT DE BARCELONA

Origen evolutivo del sentido de lo injusto y la venganza: estudio comparativo humanos-chimpancés

Evolutionary origins of unfairness and revenge: a comparative study between humans-chimpanzees

Nereida Bueno Guerra

Aquesta tesi doctoral està subjecta a la llicència ***Reconeixement- NoComercial – CompartirIgual 4.0. Espanya de Creative Commons***.

Esta tesis doctoral está sujeta a la licencia ***Reconocimiento - NoComercial – CompartirIgual 4.0. España de Creative* Commons**.

This doctoral thesis is licensed under the ***Creative Commons Attribution-NonCommercial-ShareAlike 4.0. Spain License***.

GOBIERNO DE ESPAÑA | MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE

Tesis doctoral 2017

Evolutionary origins of the sense of fairness and revenge: comparative study humans-chimpanzees

Origen evolutivo del sentido de lo injusto y la venganza: estudio comparativo humanos-chimpancés

Nereida Bueno Guerra

UNIVERSITAT DE BARCELONA

Directora:
Dra. Montserrat Colell
Universitat de Barcelona
Dpt. Psicología Clínica y Psicobiología

Co-director:
Dr. Josep Call
Max Plack Institute
for Evolutionary Anthropology

Max Planck Institute for Evolutionary Anthropology

Esta tesis sigue los "Criterios para la presentación de tesis doctorales" aprobados por la Comisión de Doctorado de la Facultad de Psicología de la Universidad de Barcelona el 2 de diciembre de 2012; los criterios que se establecen para la obtención de la Mención Internacional y ha sido aprobados por el Comité de Bioética (IRB00003099). El formato escogido para su desarrollo es el de tesis narrativa, aunque su contenido ha sido publicado/se encuentra en revisión en revistas internacionales.

This PhD dissertation complies with the "Criteria to present PhD dissertations" approved by the Doctoral Committee of the Faculty of Psychology at the University of Barcelona; the criteria to obtain the International Mention and it has been approved by the Committee on Bioethics. The format chosen is narrative, even though some of its content is already published/under revision.

Artículo publicado */Published paper:*

1. **Bueno-Guerra, N**., Leiva, D, Colell, M. and Call, J. (2016). Do age and sex affect strategic behavior and inequity aversion? *Journal of Experimental Child Psychology*, *150*, 285-300. doi: 10.1016/j.jecp.2016.05.011

Las imágenes de portada han sido reproducidas con consentimiento autorizado por escrito (enero 2017) cortesía de sus autores: rostro humano: **Frechdax** ®, Suiza; rostro chimpancé: **Gustavo Depaoli** ®, Argentina. Dado que se tratan de imágenes bajo copyright, está prohibido su uso y distribución para otro propósito que no sea la simple difusión pública de esta tesis.

The pictures on the cover have been reproduced after written consent from their authors: human face: Frechdax ®, Switzerland. Chimpanzee face: Gustavo Depaoli ®, Argentina. Since these photos are copyrighted, its use and distribution are prohibited for other purpose different from the public diffusion of the present dissertation.

Las investigaciones de esta tesis doctoral se han realizado en varios lugares durante los años 2013-2017. Los estudios con niños españoles se han llevado a cabo en centros escolares de Cataluña (España). Los estudios con chimpancés y niños alemanes se han llevado a cabo en el Wolfgang Köhler Primate Research Center del zoo de Leipzig (Germany), en colaboración con los Max Planck Insitute for Evolutionary Anthropology (MPI-EVA) y en el Human Cognitive and Brain Sciences (MPI-CBS).

The research of this PhD dissertation was conducted in several locations during the period 2013-2017. The studies with Spanish children were done in different schools across Catalonia (Spain). The studies with chimpanzees and German children were done in the Wolfgang Köhler Primate Research Center in the zoo of Leipzig (Germany), in collaboration with the MPI-EVA and in the MPI-CBS.

Instituciones colaboradoras por orden alfabético:
Collaborator institutions by alphabetical order:

Escola 25 de setembre – Rubí (España)
Escola Arrels Blanquerna – Badalona (España)
Escola Arrels Esperança – Badalona (España)
Escola Ramón Llull – Barcelona (España)
Escola Torre de la Llebre – Rubí (España)
Institut Escola Lloret de Mar – Lloret de Mar (España)
Max-Planck Institute for Evolutionary Anthropology – Leipzig (Alemania)
Max-Planck Institute for Human Cognitive and Bran Sciences – Leipzig (Alemania)
Wolfgang Köhler Primate Research Center – Leipzig (Alemania)

TESIS DOCTORAL

Origen evolutivo del sentido de lo injusto y la venganza: estudio comparativo humanos-chimpancés

Evolutionary origins of unfairness and revenge: a comparative study between humans-chimpanzees

Presentada por:
PhD dissertation by:
Nereida Bueno Guerra

Para obtener el título de:
To obtain the title of:
Doctora en Primatología – Mención Internacional

Bajo la dirección y codirección de:
Having as supervisors:

PhD Montserrat Colell Mimó
Profesora Titular de la Universidad de Barcelona

PhD Josep Call Balaguer
Director del Wolfgang Köhler Primate Research Center

Con la financiación de:
Under the economic support of:

Beca de Formación para el Profesorado Universitario (Ref. 12/00409)
Ministerio de Educación, Cultura y Deporte de España

Barcelona & Leipzig, 2017

Agradecimientos

Es curioso, porque cuando comento que mi tesis es sobre venganza la gente suele dar un pasito hacia atrás; o poner manos de kárate; o asentir sonriendo con cierto miedo. De hecho me encanta contar la anécdota de aquella vez que fui a una librería de Madrid y la cajera, al ver que todos los libros comprados llevan la palabra "venganza" en el título, me miró por encima de las gafas; puso su mano sobre la mía y me preguntó amablemente: "¿Todo bien, señorita?". Pero la reacción más habitual (bien lo sabe Héctor) consiste en subir las cejas y empezar a hacer muchas preguntas. Una de ellas es cómo puedo hacer eso de la venganza de manera ética, sin cometer ningún delito. La respuesta es que puedo hacerlo gracias a la colaboración de mucha, mucha gente.

Para un doctorado hace falta más que un doctorando y dos supervisores. Pero mi primer agradecimiento va para ellos, porque ellos han sido mis guías a lo largo de este proceso. Doy gracias a la doctora Montserrat Colell por confiar en mí desde aquel día de 2012 en que me planté en su despacho y, sin conocernos previamente pero avalando mi entusiasmo y capacidades, aceptó comenzar conmigo esta aventura. Ella me ayudó a organizar mi primera estancia en Leipzig, donde aprendí a trabajar con primates no humanos, y a lo largo de estos años me ha brindado su apoyo y ánimo en lo que a veces parecía una carrera sin fin. Desde entonces hasta hoy llenamos el despacho con fotos de bromas y viajes (ya incluyan ambulancias o ñúes). Además, mi participación en la asignatura que coordina, Etología y Evolución Humana, ha sido muy satisfactoria porque me ha permitido traducir mi experiencia investigadora en ejemplos docentes. Doy también las gracias al doctor Josep Call. Debo tener ciertas costumbres arraigadas, porque también me planté en su despacho cierto día (seguramente tras tamborilear en su puerta de cierta forma característica) e intenté convencerle para continuar en forma de tesis la línea de investigación que su equipo había emprendido hacía unos años en reciprocidad y juegos económicos. Tuve la suerte de que aceptó y, pese al abrumador volumen de estudiantes que coordina, no ha habido vez en que me haya sentido desamparada estos años. Estoy inmensamente agradecida a sus revisiones (que incluyen hasta signos de puntuación y corrección de nombres extranjeros) y sus consejos y ayuda en la toma de diferentes decisiones: siempre me han servido para mejorar. Trabajar con Josep se disfruta, es un orgullo y supone un continuo reto, por lo que puedo decir que con él he crecido mucho. Gracias a él también he podido colaborar en otras investigaciones ajenas a esta tesis, lo que me ha abierto nuevos horizontes intelectuales.

Además de la parte más profesional, la culminación de una tesis también supone disponer del apoyo incondicional de quienes conforman mi entorno personal. Los primeros de la lista, sin duda, son mis padres. A ellos les salió una niña rara, que eligió estudiar carreras poco frecuentes y hacer un doctorado todavía más infrecuente. Pero siempre me animaron a continuar con mis objetivos, me ayudaron a planificarlos e incluso, aunque no supieran inglés o no entendieran bien un aparato o una hipótesis, también me acompañaron en ellos (desde la visita a Leipzig a estar presente en una sesión experimental en Barcelona). Yo no sé qué haría sin ellos. La sensación de protección y seguridad que me han dado y el espíritu de trabajo que he vivido en casa (mi padre, a sus 55 años, empezó dos doctorados simultáneos) debe desde luego influir en parte de lo que soy ahora. Sé que mi madre hoy querría decirme muchas cosas, como que se siente profundamente orgullosa de mí, pero no puede. Pero yo quiero dejar por escrito que eso no es problema, porque lo ha dicho tantas otras veces, que tengo metidas sus palabras en el pecho.

Sigue ahora una larga lista de amigos que, no sé cómo, han sabido soportar mis enclaustramientos frente al ordenador, los picos de humor positivos y negativos y las preocupaciones y frustraciones varias que he vivido en cuatro años. Imagino que hoy deben estar muy contentos de que por fin todo termine, y por su paciencia y amor incondicional, tienen que aparecer en estas páginas. Me estoy refiriendo especialmente a Nuria, que ha vivido en primerísima persona estos años (en el que se incluyen visitas a la NASA y espejos dibujados); a Héctor, que siempre ha sabido hacerme feliz y sorprenderme, tenga yo los másteres que tenga y a Sergio, con quien he arreglado tantas veces el mundo que no entiendo cómo puede seguir roto. A ellos se les une un listado en el que no me detengo porque el lector viene a leer la tesis y no un libro de familia, pero que tienen nombre y apellidos: mi querida co-editora, Mayte Martínez, Miriam de los Santos, Isabelita García, Fabiola Sánchez, Alejandro Gutiérrez, Sergio España, Alejandro López, Carmen Reja, la familia Pardos y una persona de reciente incorporación, Cristina Gómez.

Termino ya esta sección con las personas que han colaborado, de forma directa o indirecta, en las investigaciones que se presentan. En primer lugar, gracias a las instituciones que me han acogido. Por una parte, al personal del Departamento de Psicología Clínica y Psicobiología de la Universidad de Barcelona, que me ha hecho sentir una más. En especial gracias a María Ángeles Jurado por la ayuda prestada para

asistir a congresos y realizar estancias y a Mari Carmen García, por resolverme los trámites. Por otra parte, gracias al personal del Department of Developmental and Comparative Psychology del Max Planck Institute for Evolutionary Anthropology y al personal del Wolfgang Köhler Primate Research Center. En especial gracias a Hanna Petschauer, quien con su paciencia y cariño me ha hecho sentir tan a gusto; a los cuidadores del zoo, sin los que la investigación sería imposible y que además la hacen divertida; a Raik por construir los aparatos y atender cada duda y a Johannes Grosmman y a Elvira Moehring, por resolverme los trámites. A estos agradecimientos incluyo los correspondientes a todos y cada uno de los centros que colaboraron para llevar a cabo el estudio del ultimátum (me refiero a Arrels Blanquerna y Esperança, Ramón Llull, Torre de la Llebre, 25 de setembre, Institut Lloret de Mar), y a las personas que me prestaron toda la ayuda y materiales necesarios (todos los profes y mis personas de contacto: Laura Monsonis, Yolanda, Elsa del Hort, Joana Llorens, Montse Julià, Núria Chipell, Manel Lombardía).

Tras los centros, mis últimas gracias van para los coautores. Eso significa agradecer enormemente a Natacha Mendes, Nikolaus Steinbeis y Tania Singer por la invitación a participar con ellos en el estudio de Schadenfreude y por el increíble tiempo y esfuerzo invertido en él. A ellos sumo mi agradecimiento a todos los estudiantes que colaboraron durante el experimento. Querido David Leiva, mi ínclito estadístico y tertuliano, muchas gracias. También a ti, África de las Heras, por ayudar con la recogida de datos (pellet a pellet, uva a uva) y a ti, Christoph Völter, por todo lo que me has enseñado y los análisis complejos que has hecho.

Sospecho que me quedo algunos nombres por mencionar, por lo que pido disculpas de antemano si hay ausencias. Y si las hay, os dedico especialmente a vosotros lo que viene a continuación, que al fin y al cabo es una historia de venganza.

Barcelona, marzo de 2017.

This PhD dissertation has been possible thanks to the funding provided by the Beca de Formación de Profesorado Universitario (FPU12/00409) granted by the MInisterio de Educación y Cultura during the period 2013-2017.

"La venganza representa un rompecabezas: ¿qué es lo que se gana?
La acción está hecha, el daño infligido; no se puede prevenir o deshacer.
¿O no hay ganancia para el individuo
y el deseo de venganza no es sino un impulso heredado?"
(Frijda, 1994, p.272)

"*Vengeance represents a puzzle: what is it that is gained?*
The deed is done, the harm inflicted; it cannot be prevented or undone.
Or is there no gain for the individual
and is the desire for revenge nothing but an inherited impulse?"

"Para que yo me llame Ángel González,
para que mi ser pese sobre el suelo,
fue necesario un ancho espacio
y un largo tiempo:
hombres de todo el mar y toda tierra,
fértiles vientres de mujer, y cuerpos
y más cuerpos, fundiéndose incesantes
en otro cuerpo nuevo.
Solsticios y equinoccios alumbraron
con su cambiante luz, su vario cielo,
el viaje milenario de mi carne
trepando por los siglos y los huesos."[1]
(Ángel González, 1956).

"Before I could call myself Ángel González,
before the earth could support the weight of my body,
a long time
and a great space were necessary:
men from all the seas and all the lands,
fertile wombs of women, and bodies
and more bodies, incessantly fusing
into another new body.
Solstices and equinoxes illuminated
with their changing lights, and variegated skies,
the millenary trip of my flesh
as it climbed over centuries and bones."

[1] The first time I read this poem, many years ago, I promised myself that it would appear in the epigraph of my dissertation. This poem is a tribute to evolution, to all living beings that have preceded us. If my research is aimed to dig into our evolutionary origins, the greatest of gratitude must go to our ancestors, because with their ways of living and the decisions they took, we are what we are now.

Contents

Preface

Preface

This PhD dissertation is the result of four years of intense work in Spain (University of Barcelona), supervised by PhD Montserrat Colell, and in Germany (Max Planck for Evolutionary Anthropology, MPI-EVA, and Wolfgang Köhler Primate Research Center, WKPRC), supervised by PhD Josep Call. The first idea was to study revenge through two studies: one classical experiment with children from Barcelona and chimpanzees from Leipzig (ultimatum game) plus one novel experiment with children (Barcelona). The ultimatum game is a reciprocal procedure in which one subject decides how to split some rewards with a partner, who in turn can decide whether accepting the offer (leading to payoff distribution) or rejecting it (leading to zero outcome for both). This is considered a nice way to test retaliation into the lab since it allows the opportunity to recreate unfair scenarios and paybacks. By conducting it with children and chimpanzees we could provide an evolutionary image of how revenge evolved. Our second experiment consisted of a computer-task in which some relevant subject's virtual object was eliminated by another individual and the victim could decide how to respond to the responsible of the deletion. The problem with this experiment was that we only tested children, therefore the comparative and evolutionary perspective was missing.

In order to collect the data for these experiments my permanent residence was in Spain and I travelled twice to Germany (in 2014 and 2015) in the form of two research stays of 3 months each[3]. However, those stays were not my first time at the MPI-EVA and WKPRC. My first stay had been in 2012, when I went to training as a research assistant in Primatology. During this training I participated in some studies. One of them was called "Schadenfreude", which is now one of the core studies of the present dissertation. What happened with my participation in that experiment that changed the course of my initial PhD plan?The Schadenfreude experiment, originally conceived by the research group comprised by Natacha Mendes, Niko Steinbeis, Tania Singer and Josep Call, was aimed to test whether children and chimpanzees, similarly to human adults, actively (i.e. incurring a cost) seek for watching how someone who treated them badly in the past (a nasty person) is punished. The results were promising, as both children[4] and

[3] The stays were funded by Short-term grants from the Spanish Ministry of Education, Culture and Sports.
[4] The children data was collected by Christine Brenner, Katharina Mueller, Charlotte Hoecker, and Jessica Buergel at the Max-Planck Institute for Human Cognitive and Brain Sciences (MPI-CBS). I collected most of the chimpanzee data but I need to thank the assistance of Matthias Allritz, Vera Ehrich, Kerstin Esau, Elisa Felsche, Johannes Grossmann, Susan Hunger, Saskia Lorenz, Julia Steinhardt, Katrin

chimpanzees showed some similar patterns of behavior: both species were willing to incur costs to watch the antisocial punishment However, it was unclear whether chimpanzees really understood the nastiness of the experimenters and, consequently, whether their reactions could be really considered Schadenfreude. Therefore when the results were submitted to a peer-reviewed journal, one reviewer suggested to adding a new condition in which the nasty experimenter did not address her actions to the chimpanzee but to a conspecific (i.e. adding an "indirect condition"). As it is very well proven that chimpanzees do not consistently incur costs in actions related to third-parties, the inclusion of this condition could help to better interpret the results. In case the chimpanzees experienced Schadenfreude for real, they would need to show some reaction in direct conditions and no reaction in indirect conditions. In case chimpanzees did not have understood the task nor experience Schadenfreude, then they should behave the same in both conditions.

Dr. Josep Call invited me to collect the data for this new condition in a new 3-months research stay (in 2016) and kindly included me as a co-author of the manuscript. As a result from that stay, we decided to include the Schadenfreude experiment in the present dissertation for three reasons: it was certainly related to revenge (there is a victim, there is an offender and there is some payback that can be enjoyed in case the victim was the subject itself); it was conducted with children and chimpanzees (the evolutionary perspective was obvious); and it could add the emotional part to the behavioral response that we already had. Therefore, what the reader is about to find is a collection of three studies: ultimatum game with children, ultimatum game with chimpanzees[5] (revenge response) and Schadenfreude with children and chimpanzees (revenge emotion).

Finally, due to time constraints related to the revision process of the manuscripts, I have been eventually forced to present this dissertation following the classical format instead of being a compilation of publications. Nevertheless, it is worth to note that the ultimatum game experiment with children has been already published (Bueno-Guerra et al, 2016) and the Schadenfreude experiment has passed through three revisions now. This is why some parts of the present dissertation might appear in the already existent/about to come publications.

Schumann, Katja Waldherr and Katharina Wenig during the training phase or during the test phase (playing the corresponding nice/nasty role).

[5] I consider the ultimatum experiments separately because their methodology was very different (one-shot in the case of children; iterated in the case of chimpanzees: see Introduction).

Resumen y Objetivos

Summary and Objectives

Resumen (versión en castellano):

En la presente tesis se aborda el estudio evolutivo del castigo. El castigo se puede entender como "functional punishment" cuando en un sistema repetido de interacciones acaba cumpliendo una función, como es por ejemplo la de amedrentar al causante del conflicto y forzar su cooperación (Clutton-Brock & Parker, 1995; Jensen, 2010). Otro tipo de castigo es el "functional spite" cuando tras un único conflicto el agente víctima decide responder el evento mediante un acto costoso que en principio no parece generarle beneficio alguno (aunque de manera indirecta o a efectos evolutivos sí los produzca). A este tipo de castigo se le conoce como "venganza" (Frijda, 1994). Para estudiar en el laboratorio estos dos tipos de respuesta se suelen utilizar los juegos económicos, tales como el ultimátum o el dictator (Güth, Schmittberger, & Schwarze, 1982). Ambos se basan en el reparto entre dos sujetos de unas recompensas. Un sujeto (oferente) escoge la forma de reparto con el otro sujeto (receptor). En el dictator, el receptor debe aceptar cualquier reparto, mientras que en el ultimátum, tiene poder para aceptarlo o rechazarlo. En caso de rechazo, ambos sujetos no obtienen nada. Para estudiar *functional punishment* interesa realizar de manera repetida el ultimátum para comprobar si los receptores adoptan el rechazo (castigo) como manera de provocar que el oferente haga mejores ofertas en el futuro. Si el oferente comprende esta dinámica, debería entonces cambiar su comportamiento en sucesivas interacciones. Dado que en el dictator no hay oportunidad de réplica por parte del receptor, el oferente no tendría por qué cambiar su conducta, que podría ser egoísta. Por tanto, para estudiar la filogenia del functional punishment cabe ver si los chimpancés rechazan ofertas injustas de comida, mientras que para comprobar su función cabe ver si los chimpancés oferentes cambian su comportamiento en el ultimátum pero no en el dictator.

Aunque el ultimátum ya se ha llevado a cabo en chimpancés con anterioridad (Jensen, Call, & Tomasello, 2007a; Proctor, Williamson, de Waal, & Brosnan, 2013a; Riedl, Jensen, Call, & Tomasello, 2012), ha habido diversas preocupaciones metodológicas que no permiten concluir con certeza sobre la conducta de los chimpancés: o bien no se siguieron diseños ABA (entrenamiento no social; test; post-entrenamiento no social); o no se tuvo en cuenta la poca intervención del experimentador para evitar interpretaciones alternativas; o no se añadió un dictator game con el que poder llevar a cabo la comparación de respuestas. En nuestro experimento contamos con 4 díadas de chimpancés que realizaron: entrenamiento, ultimátum, post-entrenamiento, dictator,

post-entrenamiento (variando el orden de los tests entre diadas para juzgar efectos de orden). Además, el experimentador no intervino durante el proceso de decisión.

Por otra parte, para estudiar *functional spite* se puede emplear el mismo juego, el ultimátum, pero jugado en una única ocasión ("one-shot"). Que se juegue solo una vez implica que rechazar, es decir, ejecutar el castigo, suponga un coste para quien lo ejecuta pero no le suponga un futuro beneficio ya que no habrá una futura interacción con ese mismo compañero. A través de los rechazos podemos saber si los niños realizan acciones de venganza o no, mientras que a través de las ofertas podemos comprobar si los oferentes son capaces de prever esos rechazos a priori y proponer ofertas equilibradas o justas con respecto a su compañero. El ultimátum ya se ha realizado con anterioridad en niños (Gummerum & Chu, 2014; Güroğlu, van den Bos, & Crone, 2009; Harbaugh, Krause, & Liday, 2003; Sutter, 2007; Wittig, Jensen, & Tomasello, 2013), sin embargo la parte metodológica ha sido muy confusa porque utilizaron el *strategic method* (simular una interacción) en vez del *direct-response* (interacción real); o usaron sistemas de conversión de recompensas (las que estaban en juego no eran las recompensas reales que se dividían ambos jugadores), complicando por tanto la interpretación de los resultados. En nuestro experimento, aproximadamente 500 niños de 6 y 10 años jugaron el juego de manera real, sin reglas de conversión y de forma anónima. Analizamos si la generosidad estratégica de las ofertas y la tasa de rechazos variaban con la edad o el sexo.

Por último, el sentimiento de disfrute asociado con observar cómo se ejecuta un castigo sobre un agente que ha causado un daño previo se denomina "Schadenfreude", una palabra derivada del alemán que significa "alegría por el dolor". Los teóricos todavía discuten si se trata de un sentimiento anterior a la ejecución de un acto ofensivo (ergo motivador) o posterior (ergo consecuencial). No se ha realizado hasta la fecha ningún estudio con niños y chimpancés en esta temática, sino con adultos en situaciones de endogrupo y exogrupo (por ejemplo, ver a los rivales de un equipo de béisbol sufriendo una derrota) (Cikara & Fiske, 2013). En nuestro experimento, niños y chimpancés interaccionaron con un agente prosocial (que les daba comida/juguetes) y antisocial (que les quitaba comida/sujetes) y posteriormente un tercer agente aparecía en escena para pegar a cada uno de los agentes previos (en dos momentos temporales diferentes, con orden contrabalanceado). Para ambas especies el castigo dejaba de ser visible (bien mediante un telón, estilo teatro, o bien desplazándose a otra sala) y la variable

dependiente en la que estábamos interesados era en la respuesta para continuar viéndolo: si los niños gastaban unas monedas adquiridas previamente para destapar el telón y si los chimpancés se esforzaban por abrir una puerta que les llevaba al lugar donde se desarrollaba el castigo. En el caso de que experimentasen Schadenfreude, ambas especies deberían incurrir un coste para continuar viendo el castigo del agente antisocial pero no del prosocial. En el caso de los chimpancés, además, repetimos la misma secuencia de acontecimientos pero esta vez los agentes interaccionaban con un conespecífico, de manera que veíamos si los chimpancés consideraban lo que les sucede a terceras partes como igualmente satisfactorio para continuar viéndolo, pese a que alguna evidencia previa sugiere que tal vez los chimpancés no realicen esfuerzos en tareas de laboratorio ante las ofensas que reciben terceras partes (eg. Riedl et al., 2012).

Con estos estudios esperamos poder arrojar más luz sobre los orígenes evolutivos del castigo: su posible función para fomentar la cooperación entre individuos (estudio 1); su posible existencia como evento costoso que no produce beneficios claros (estudio 2) y sus emociones asociadas cuando no se ejecuta directamente pero puede presenciarse (estudios 3 y 4).

Objetivos (versión en castellano):

ESTUDIO 1: MINI-ULTIMATUM Y DICTATOR EN CHIMPANCÉS

1. Estudiar la filogenia del **castigo funcional** (= "functional punishment") como recurso para producir interacciones cooperativas.
   1.1. Explorar la existencia de rechazos en un juego de distribución de comida iterativo entre dos chimpancés cuando la oferta recibida produce un reparto de recompensas desigual.
   1.2. Analizar si esos rechazos producen en el compañero un cambio hacia la igualdad (=cooperación) en futuras ofertas de comida.
   1.3. Explorar si la generosidad de los chimpancés que ofrecen comida cambia según la capacidad que tenga el compañero para castigarles.
   1.4. Explorar si los gestos/comportamientos que el compañero hace antes del reparto influyen en la generosidad de quien reparte.
   1.5. Analizar la existencia de reciprocidad positiva/negativa entre chimpancés.

ESTUDIO 2: MINI-ULTIMÁTUM EN NIÑOS

2. Estudiar la ontogenia de la **venganza** (= "functional spite")
   2.1. Explorar la existencia de rechazos cuando niños de 6 y 10 años reciben repartos desiguales en un juego de distribución de pegatinas y saben que no habrá futuras interacciones con su compañero anónimo (= "one-shot game")
   2.2. Analizar si esos rechazos varían en función de la intención de desigualdad que tenía el compañero que hizo el reparto.
   2.3. Explorar si los niños de 6 y 10 años que ofrecen pegatinas anticipan rechazos potenciales en sus compañeros y ofrecen repartos generosos a priori (= "strategic behavior").
   2.4. Analizar si la edad y el sexo son factores determinantes para la ejecución de rechazos o las ofertas a priori generosas.

ESTUDIOS 3 Y 4: SCHADENFREUDE EN NIÑOS Y CHIMPANCÉS

3. Estudiar la filogenia y ontogenia del placer que se siente al presenciar el castigo de un agente que previamente ha provocado un daño (= "Schadenfreude")

3.1. Analizar si chimpancés y niños incurren gastos para continuar viendo el castigo de un agente que previamente les provocó un daño antes que para continuar viendo el de un agente que fue amable con ellos.

3.2. Explorar si el comportamiento de los chimpancés cambia según si los agentes interaccionaron directamente con ellos o con otro sujeto de su misma especie (= "second-party vs. third-party").

3.3. Registrar los comportamientos emocionales que niños y chimpancés ejecutan cuando presencian los castigos y analizar si varían en función de la valencia del agente.

Summary (English version):

This thesis deals with the evolutionary study of punishment. Punishment can be understood as "functional punishment" when in a repeated system of interactions it ends up fulfilling a function, such as threatening a free-rider and forcing his/her cooperation (Clutton-Brock & Parker, 1995; Jensen, 2010). Another type of punishment is "functional spite" when after a single interaction the victim decides to respond to the offender in a costly way even when this does not seem to generate any direct benefit in return (the benefit can be indirectly or evolutionarily achieved). This type of punishment is known as "revenge"" (Frijda, 1994). Economic games help to study these types of punishment in the lab, specifically, the ultimatum and the dictator games (Güth, Schmittberger, & Schwarze, 1982). Both games consist of distributing resources between two subjects. One subject (proposer) chooses how to divide the rewards with his partner (recipient). In the dictator game the recipient must accept any distribution, while in the ultimatum game, the recipient can decide whether accepting or rejecting the offer. In case of rejection, both subjects get nothing. In order to study functional punishment, it is interesting to run the ultimatum game iteratively to see whether recipients adopt rejection (punishment) as a way to promote a change in the proposer towards the selection of more generous offers in the future. If the proposer understands this dynamic, then he should change his behavior in successive interactions. Given that in the dictator game the recipient is passive, the proposer would not necessarily need to change his behavior, which could be selfish from the beginning to the end of the game. Therefore, to study the phylogeny of functional punishment, we should check whether chimpanzees reject unfair food offers, and whether those rejections produce a change in the proposer's behavior when playing the ultimatum but not the dictator game.

Despite chimpanzees have been previously tested in the ultimatum game (Jensen, Call, & Tomasello, 2007a; Proctor, Williamson, de Waal, & Brosnan, 2013a; Riedl, Jensen, Call, & Tomasello, 2012), there have been several methodological shortcomings that preclude concluding about how chimpanzees bargain: either no ABA (non-social training, non-social post-training) designs were followed; or the experimenter's intervention was excessive; or there was no dictator game to compare responses with. In our Study 1 we had 4 chimpanzee dyads playing: non-social training, ultimatum, non-social post-training, dictator, non-social post-training (counterbalancing the order of

tests between dyads to avoid order effects). In addition, the experimenter did not intervene during the decision process.

On the other hand, the study functional spite can benefit from using the ultimatum game but as long as it is played on a single interaction ("one-shot"). Rejecting in one-shot games supposes a cost with no clear future benefit (since there will be no future interaction with the same partner). Through rejections we can get to know whether children take revenge, while through offers we can explore whether children are able to anticipate rejections and propose generous offers from the very beggining. The ultimatum has already been conducted with children (Gummerum & Chu, 2014; Güroğlu, van den Bos, & Crone, 2009; Harbaugh, Krause, & Liday, 2003; Sutter, 2007; Wittig, Jensen, & Tomasello, 2013). However, some methodological shortcomings have produced confounded data. For example, most of the studies used the strategic method instead of the direct-response. Additionally, most of them used reward conversion systems rather than a "what you see is what you get" system, thereby complicating the interpretation of the results. In our experiment, approximately 500 children aged 6 and 10 played anonymously the game in a real scenario (direct-response method), following the what you is what you get system. We analyzed whether sex and age had an influence on strategic behavior and on the rate of rejections.

Finally, the enjoyment associated to witnessing the punishment of a previously harmful agent is called "Schadenfreude", a German compound word for "joy for others' misfortune". Theorists still argue whether Schadenfreude is experienced prior to the execution of an offensive act (ergo it works as a motivator) or it is experienced after the punishment (ergo it is an emotional consequence). To date there are no studies with children and chimpanzees on this subject but with adults comparing the pleasure felt after witnessing endogroup's and exogroup's misfortunes (for example, watching plyers from the supported/rival baseball team losing a game) (Cikara & Fiske, 2013). In our experiment, children and chimpanzees interacted with a prosocial agent (gave them food/toys) and with an antisocial agent (who took food/toys away from them). After that, a third agent came in to hit each of the previous agents (at two different separate times, order counterbalanced across subjects). The punishment became out of view (either through a theater curtain or through moving to a different room). The dependent variable we were interested in was the costly response to make the punishment visible again: coins expenditure (children) to uncover the theatre curtain again; physical effort

to open a heavy sliding door to get access to the room where the punishment was being developed. Should both species experience Schadenfreude, they should incur costs to continue watching the punishment of the antisocial (but not the prosocial) agent. In the case of chimpanzees we repeated the same sequence of events but this time the agents interacted with a conspecific. Thus, we explore whether chimpanzees care about third-parties and find equally satisfying to continue watching the (indirect) antisocial agent as well. We were not very optimistic about finding indirect Schadenfreude in chimpanzees since some previous lab evidence suggests that chimpanzees may not incur any costs to react against third-party offenders (eg Riedl et al., 2012).

With these studies we hope to shed more light on the evolutionary origins of punishment: its possible key role in promoting cooperation between individuals (study 1); its possible existence as a costly event that does not seem to produce clear benefits (study 2) and its emotional byproducts when the punishment cannot be directly executed but witnessed (studies 3 and 4).

Objectives (English version):

STUDY 1: MINI-ULTIMATUM AND DICTATOR GAMES IN CHIMPANZEES

1. Study the phylogeny of functional punishment as a way to produce cooperative interactions.
   1.1. Explore the existence of rejections in an iterative food distribution game between two chimpanzees when the offer received produces an uneven rewards distribution.
   1.2. Analyze whether these rejections produce a change towards equality (=cooperation) in the proposer across future food offerings.
   1.3. Explore whether the chimpanzee offering food's generosity changes depending on the partner's capacity to punish him/her.
   1.4. Explore whether the communicative attempts that the partner executes before the distribution have an influence over the proposer's generosity.
   1.5. Analyze whether positive / negative reciprocity occurs.

STUDY 2: MINI-ULTIMATUM GAME IN CHILDREN

2. Study the ontogeny of revenge (= "functional spite").
   2.1. Explore the existence of rejections when 6- and 10-year-olds receive unequal distributions in a one-shot sticker distribution game.
   2.2. Analyze whether these rejections vary according to the partner's intention to be generous.
   2.3. Explore whether 6- and 10-year-old proposers anticipate potential rejections and consequently choose generous distributions (= "strategic behavior").
   2.4. Analyze whether age and sex are determining factors for the execution of rejections or strategic offers.

STUDIES 3 AND 4: SCHADENFREUDE IN CHILDREN AND CHIMPANZEES

3. Study the phylogeny and ontogeny of the pleasure delivered from witnessing the punishment of an agent who was previously harmful (= "Schadenfreude").
   3.1. Analyze whether chimpanzees and children incur costs to continue watching the punishment of an antisocial agent rather than the punishment of a prosocial agent.
   3.2. Explore whether chimpanzees' behavior changes according to whether the agents interacted directly or indirectly with them (= "second-party vs. third-party")
   3.3. Code children' and chimpanzees' emotional behavior during the witnessing of the punishments and analyze whether it varies depending on the antisocial/prosocial nature of the agents.

Introduction

Introduction

It's Wednesday in Oklahoma. You are enjoying a day free at home, lazily cleaning the living room while listening to the radio. Your beloved husband/wife is at his/her workplace, a nine-floor federal building located in 200 N.W. 5th Street. The minute hand on the wall clock arrives at 9.02 a.m. and suddenly there is a shocking and fatal deafening noise outside. Traffic stops; the murmur of people in the cafes as well. A shuddering atmosphere expects the worst. Even though you had never heard an explosion before, you instantly realize that that was a bomb. The radio talker announces with a serious voice the location of the terrorist attack. It's 200 N.W. 5th Street. On that April 19th 1995 things happen disorderly. After hours of expectation and agitation without a relief phone call, you learn that your beloved one is dead, together with 167 more people. In the meanwhile, Timothy McVeigh, the perpetrator, and Terry Nichols, his co-conspirator, are caught by the police. McVeigh is sentenced to death by lethal injection and Nichols is convicted with life imprisonment without parole. In 2001, you receive a letter from the Government. It is an invitation scheduled for June 11th to witness how McVeigh is executed. Would you confirm your attendance?

There is no need to feel guilty or shame if your response is affirmative. Feeling pleasure at someone else's misfortune is called "Schadenfreude"[6] (Heider, 1958) (a borrowed German compound word formed by Schaden (damage, harm) and Freude (joy), hence "the joy of harm"). Revenge, vengeance or payback (terms that I will use indistinctly through this dissertation[7]) are defined as a motivated desire to punish another individual for his/her offence. It is true that contrary to gratitude revenge and Schadenfreude seems pejorative, immoral and socially condemned in our society (Murphy, 2011; Summerfield, 2002). Besides, during the twentieth century revenge was even considered unhealthy and pathological (Daniels, 1969; Horney, 1948; Socarides, 1977) because it

[6] I will keep the "S" in capital letter through this dissertation despite in English it is not capitalized as a respectful tribute to the Germanic origins of the word.

[7] In a personal communication (2012), Keith Jensen, a psychologist that has studied revenge in apes, told me that he could not provide a clear-cut distinction between the words "revenge" and "vengeance" as a native English speaker. However, for subtle differences between the so-called three R's (retaliation, revenge and redirect aggression), please refer to Barash and Lipton (2011, pp. 4-5): "When one being hurts another, several things may happen. Sometimes, the pain is immediately reflected back onto the perpetrator. This is retaliation. It is prompt and straightforward. (…) Then there is revenge. (…) The response is delayed –often for a long while and with much prior contemplation. (…) The strangest form of payback (…) is redirected aggression (…) –the targeting of an innocent bystander in response to one's own pain and injury." Despite these differences, the scientific literature on revenge does not follow a homogeneous standard definition for all these words. Therefore, just to avoid word repetition through the writing, I will use revenge, vengeance and payback indistinctly.

conferred the individual an antisocial, violent or aggressive reputation. However, many authors are recently claiming that "instead of human nature's being either fundamentally brutish or fundamentally noble, it is both" (de Waal, 1996, p. 5), proving that revenge has endured since the earliest stages of civilization (Jacoby, 1983) and it is as regular and enrooted as other more socially acceptable reactions, such as forgiveness (McCullough, 2008; McCullough, Kurzban, & Tabak, 2012). If that is true, we should be able to find ample evidence across human History of these two phenomena; some evolutionary explanation for its persistence and, if both phenomena are adaptive, we should also find some evidence of the underlying mechanisms of revenge and Schadenfreude in other species, more concretely in our closest living relatives, the chimpanzees. Therefore, this is the structure that I will follow across the present dissertation.

First, in "Revenge and Schadenfreude across Human History" section I examine some historical examples about the existence of revenge and Schadenfreude across different cultures and continents. My objective there is to prove how enrooted these responses in humans are. Second, in the following sections "Punishment: types and functions"; "Revenge: the punishment with an apparent function" and "Schadenfreude: the enjoyment of punishment", I provide a general explanation about punishment, define both terms, discuss their functions and comment some previous studies. My objective there is to show why these responses are so adaptive that they have been naturally selected and how they have been studied in the lab. Finally, in "Our studies section" I introduce the experimental studies with children and chimpanzees that comprise this dissertation. My objective there is to present how we decided to explore whether revenge and Schadenfreude share some phylogenetic pathway.

Revenge and Schadenfreude across Human History

History and Literature provide unempirical evidence of the fact that revenge and Schadenfreude have been (and still are) widely spread in humans. From the Punic wars to the Treaty of Versailles from the Alexander the Great's expeditions against Persia to the forty-seven rōnins[8], from the Tutsis and Hutus to Israelis and Palestinians, vengeance has fueled human History across different countries and ages (Bueno-Guerra,

[8] The 47 rōnins (the Japanese name for samurais without a master) sought revenge upon their master after he was forced to kill himself for having offended a public servant. This historical event is one of the most ancient Japanese vendettas that has been traditionally played with Japanese puppets. The story can be found in Mitford's book (1871), recently republished by Dover Publications in 2005.

2012; Ugidos, 2015). Contemporary History is no exception. One of the most well-known retaliatory events to which the reader is likely to be familiar with is the assassination of Osama Bin Laden in 2014, considered as a vindictive response of the United States after the 9/11 attacks (Bowden, 2012). In fact, the Americans felt so negative emotions when recalling that day (Lambert et al., 2010) that the narration of the assassination has been used in current research as a stimulus to elicit revenge feelings in experimental subjects (Lambert, Peak, Eadeh, & Schott, 2014). Revenge is also a recurrent headline in the recent news worldwide since different politicians vow for revenge within their international affairs outlook. "Revenge is unavoidable" (17th November 2015)[9] said Putin, the President of Russia, after a Russian jet was downed by Syrian rebels. "Divine revenge will seize them" (3rd January 2016)[10] warned the Iranian supreme Ayatollah leader after the execution of one of their prominent clerics. "Revenge will come" (12th December, 2016)[11] promised the Interior Minister Suleyman Soylu after Istanbul bombings.

The joyful experience of watching others suffering has also a long historical tradition. During the Roman Empire, the killing of a gladiator in the coliseums was cheerfully announced and celebrated[12]. In the Middle Ages there was the trend to represent tortures, "as if violent time bred enjoyment of violence" (Tuchman, 2011, p. 312). Moreover, public executions were enjoyed in Spain until the end of the XIX century: the audience would boisterously crowd the market squares (Arenal, 1867) arriving from different villages in the nearby just to watch the execution (Sueiro, 1974). There were food and drinking stalls (Lucea Ayala, 2004) and people would fight for a place in the first rows, happily enjoying the act of killing and already hoarse after too much enthusiastic shouting (Sueiro, 1974) with "satisfaction" (Montes Luengo, 1897). Similar joyful reactions aroused in the audience while watching how witches were burnt across Europe until the XVIII century (Russell & Alexander, 2007). Even more recently, that joy could have been shared by the attendees to the execution mentioned in the

[9] CNN. (n.d.). Vladimir Putin: 'Revenge is unavoidable' [Video]. Retrieved from: http://edition.cnn.com/videos/world/2015/11/17/putin-syria-airstrike-metrojet-matthew-chance-lkl-newsroom.cnn

[10] BBC News. (2016, January 3). Iran: Saudis face 'divine revenge' for executing al-Nimr [Video and report]. Retrieved from http://www.bbc.com/news/world-middle-east-35216694

[11] Gurcan, M. (2016, December, 14). Turkish state 'out for revenge' after Istanbul bombings [Report]. Retrieved from http://www.al-monitor.com/pulse/ru/contents/articles/originals/2016/12/turkey-promises-revenge-after-deadly-bombings.html

[12] Interestingly, in his poem "The gladiator" Lord Byron uses the metaphor "butcher'd to make a Roman holiday" as a synonym of the celebration that would be held after watching the suffering of the gladiator.

introductory story of this dissertation. That story is based on real facts, concretely the Oklahoma City Bombing. This terrorist attack, the worst second in number of victims in the U.S. after the 9/11[13], is especially relevant to the topic of revenge and Schadenfreude because it was the first time in the American History that an execution was broadcasted to fulfill the emotional needs of the victims (Madeira, 2012, p. 184-85):

> "In January 2001, family members and survivors were asked to indicate whether they would be willing to attend the execution live in Terre Heute. These… seats would be assigned by lottery…with seven seats reserved for victims' family members, two seats reserved for survivor with physical injuries and one seat reserved for other survivors…
>
> As the execution date drew near, numerous Oklahoma City bombing survivors and victims' families… sought to persuade… to arrange for a closed-circuit broadcast of the execution… They framed the struggle to televise McVeigh's execution in terms of their need to witness "justice"".

It is interesting to notice what the definition of "victim" was in order to be allowed to attending the execution in the execution chamber: those directly harmed by the bombing attack, prioritized in order of relevant damage. This gives an idea about who were considered the deserving agents of revenge and Schadenfreude: people who had received a previous harm. However, that definition did not satisfy all. Hence, a group of representatives asked whether the witnessing of the execution could be permitted not only for those who lived the event directly (suffered the harm in first-person, i.e. survivor, relative) but also for those who lived it indirectly (witnessed how a third-party suffered, i.e. rescuers, witnesses). Importantly, revenge and Schadenfreude were proving not to be limited to those directly affected but also to those indirectly affected, namely the bystanders of harm. The pragmatic issue was that with this request the number of potential attendees greatly exceeded the limit permitted in the execution chamber. Therefore, the solution was to set an *ad hoc* place to broadcast the execution on closed-circuit television, which eventually congregated more than 200 people:

[13] A complete list of the victims is provided here: http://www.webcitation.org/5wovK3hIw (Last accessed: December 31st 2016).

> "The decision to broadcast the execution on closed-circuit television had come after many people connected to the bombing had complained to the Justice Department. More than 1.000 invitations were mailed to survivors, rescuers, family members of victims and others. Although 325 people responded that they would attend, 232 arrived at a designated parking location near the local airport, some as early as 2 a.m."[14]

This very recent event in our History shows the relevance that applying punishment to wrongdoers and watching that punishment has for humans. In fact, revenge and Schadenfreude are not only present in the historical becoming but also cover the artistic production, which has produced numerous texts dealing with these topics from ancient times (Barash & Lipton, 2011). Just to mention some, the reader might be familiar with Medea, the famous Euripides' tragedy which depicted a woman killing her offspring to make her recently re-married husband suffer. The Count of Monte Cristo is also a very well-known Dumas' story of an inmate who escaped from prison to face those who incarcerated him unfairly. The Shakespearean play Hamlet portrayed a prince who kills his father's murderer. More recently, there is even a literary trend aimed to eulogize these topics. Just to name a few: "V for Vendetta" (Moore & Lloyd, 1990); "Sweet revenge (The Wicked Delights of Getting Even)" (Barreca, 1995); "The Book of Getting Even: A Novel" (Taylor, 2008); "Schadenfreude: The Little Book of Black Delights" (Lihoreau, 2011); "Whipping Boy. The Forty-year Search For My Twelve-year-old Bully" (Kurzweil, 2015). This trend is extensible to the film industry, which has produced a vast array of vindictive or painful-enjoyable stories as well, such as the Korean "Trilogy of Vengeance", the Japanese manga "Lone Wolf and Cub", the American bloody "Kill Bill". There are also explicitly Schadenfreude motivated productions, such as "Jackass", which showed how people participated in different extremely painful activities while their friends laughed at them or the Japanese imported TV show "Takeshi's Castle" ("Humor Amarillo" in Spanish), where contestants fell and crashed repeatedly. This tribute to vengeance and entertaining other people's suffering has even reached the possibility to invite the spectators to vote for which punishments they mostly preferred to watch being applied to the wrongdoer characters in the movie

[14] Yardley, J. (2001, June 12). "The McVeigh Execution" [Report]. Retrieved from: http://www.nytimes.com/2001/06/12/us/the-mcveigh-execution-oklahoma-city-execution-on-tv-brings-little-solace.html

they are watching. The "interfilm" (a short name for *interacting films*) "Mr. Payback" provided the spectators with remote controls so that the plot of the film projected changed according to the majority of the votes from the audience:

> "Mr. Payback (Billy Warlock) is "a hero… programmed to seek outrageous justice for his clients against those who have done them wrong." An opening sequence introduces viewers to… Mr. Payback [who] states… his mission, to rectify the evils of society. Then viewers choose from three plot lines... In one, a woman has been sexually harassed; in another, a black man has been fired from his job and cheated out of a large sum of money; in the third a Hispanic man has been discriminated against and called a "taco eater".
>
> In each case… it's payback time. Mr. Payback gets on the job, and… exacts revenge upon the villain responsible for his client's humiliation. The final scene always involves the wronged individual confronting the wrongdoer… and meting out to him his just desserts, the exact nature of which are determined by audience participation.
>
> Viewers push buttons on pistol grips mounted on the armrests of their seats; graphics indicate when it's time to vote and what the choices are… The idea is to push your button of choice not just once-which one would think would be all that was strictly necessary-but as many times and as fast as possible. This is clearly designed to foster an atmosphere of rowdy and cheerful competition, especially as audience members are encouraged to shout at one another."[15]

With all these references, it is conceivable to say that revenge has contributed to a relevant part of human History. Besides, it seems that when punishment has been made available to audiences, both victims (those directly affected by a previous offense) and bystanders (those indirectly affected) have enjoyed (i.e. felt Schadenfreude) watching others suffer, especially if the suffering individual had committed some previous offense. However, despite their insidious presence across time, the logic and the usefulness underlying revenge and Schadenfreude might not be that easy to determine.

[15] Werner, E. L. (1995, February 23). Are we having fun yet? [Report]. Retrieved from: http://www.thecrimson.com/article/1995/2/23/are-we-having-fun-yet-pbmr/

Punishment: functions and types

I am aware that I have started talking about revenge and Schadenfreude without providing clear-cut definitions that circumscribed the kind of situations that we can discuss. But it was on purpose. We all as humans have a folk conception of what both terms are. For example, in a school I visited during this research I asked a group of more than thirty 10-year-olds to write what revenge was for them. Even though in their early life it was unlikely to have accumulated many social conflicts, or even heard the term many times before, the average response was: "*To do something wrong to someone that did something wrong to you before*". It is conceivable that the majority of us would agree with this definition. Revenge is a response to a previous harm. It is also very common to have experienced joy while seeing someone that we do not like failing or having bad luck. Schadenfreude is being happy at other's misfortune. Therefore, what revenge and Schadenfreude mainly have in common is punishment. Revenge is the punishment applied to an offender (i.e. the response) whereas Schadenfreude is the enjoyment of the application of that punishment (i.e. the emotional byproduct). As a result, when looking for a theoretical framework of both, we should start with punishment.

Punishment is a costly response executed after some previous negative event (Clutton-Brock & Parker, 1995; Jensen, 2010). Therefore we can refer to punishment as "negative reciprocity" as well, since a negative event is followed by another negative event (Heider, 1958). It is not necessary that the punisher suffers the negative event in order to become a punisher. We will refer to second-party punishment when the punisher was the previous victim, and to third-party punishment when the punisher and victim do not coincide. Usually, in third-party situations the punisher is a bystander of some negative event who voluntarily decides to act against the offender. This distinction has also implications on the functions of punishment. The evolutionary biologists distinguish between different types of punishment depending on the benefits that the punisher obtains and consequently depending on its likelihood to evolve: if the punisher obtains some direct benefit, the punishment is selfish and it is understandable that it had evolved because it increases the punisher's fitness. This example of negative reciprocity is called "functional punishment". For example, an individual punishes someone that has previously attacked him and thereby prevents to be attacked again in the future. However, if the recipient of the benefits derived from the punishment is the group, not

the punisher, then the punishment is still functional but it is most commonly denominated as "altruistic punishment". The adjective altruistic is used because altruism refers to costly acts to the actor's fitness that produce a fitness benefit for another individual (Hamilton, 1964). Thus, altruistic punishment is likely to evolve because it helps to increase both related and unrelated individuals' fitness, who benefit from the deterrence of a non-cooperator. Indeed, altruistic punishment has been argued to be crucial for human cooperation (Fehr & Fischbacher, 2003) and it is very rare among non-human animals (Leimar & Hammerstein, 2010; Melis & Semmann, 2010).

The revolutionary idea for the study of punishment came with the inclusion of reciprocity. In a world of repeated interactions initial costly actions could be reimbursed in the future (although this does not mean that individuals calculate reciprocity (Noë & Hammerstein, 1994)). In his seminal paper, Trivers (1971) first showed this with costly positive reciprocity and coined the expression "reciprocal altruism". Reciprocity was a natural selected mechanism compatible with selfish (but willing to incur costs) individuals because it helped in building systems of mutual cooperators. A decade later, in another seminal paper, Axelrod and Hamilton (1981) expanded the idea of reciprocal altruism to functional punishment. To do that, they used game theory. Game theory is an economic framework where individuals and resources are placed together and researchers explore how individuals interact under certain rules (ultimatum game, dictator game, Prisoner's dilemma, public goods are examples of game theory experiments). Concretely, Axelrod and Hamilton used an iterated Prisoner's dilemma. Briefly, the Prisoner's dilemma is a game where individuals face the decision to cooperate with or to cheat each other, what gives a resultant matrix of gaining and loses: if both cooperate, they gain a similar low payoff (reward for mutual cooperation); if both cheat, they experience loses (punishment for mutual defection); if one cheats and the other cooperates, the cheater obtains a higher payoff than the cooperator (Rapoport & Chammah, 1965). This is a very useful paradigm to reproduce low-scales simulations of evolutionary facts, since resources can be understood as fitness and repeated interactions as time scales. With such scenario, cheating was tempting. Thereby the inclusion of punishment for cheating allowed to show the "initiation of cooperation from a previously asocial state" (Axelrod & Hamilton, 1981, p. 1391) and not only from the pure positive reciprocity perspective argued by Trivers. Their main result was that the best evolutionary engine for cooperation was the tit-for-tat strategy. This strategy

was simply based in echoing the previous partner's action: cooperate after cooperation; defeat after defection. Human cooperation now is explained through this reunion between positive and negative reciprocity. If this robust reciprocal rule is consistently enacted, it is usually referred as "strong reciprocity"[16] (Fehr, Fischbacher, & Gächter, 2002; Gintis, 2000; but see Guala, 2012 for differentiation between weak and strong reciprocity and de Waal & Brosnan, 2006 for the inclusion of time delay in the definition of strong reciprocity). In particular, defeating after defection is interesting to explain punishment because it gave path to the understanding of its function: costly punishment was useful because a taste of your own medicine proved to increase cooperation and deter future cheating. In this sense, the useful thing of punishment was "that reduces the fitness of the instigator and discourages or prevents him or her from repeating the initial action"[17] (Clutton-Brock & Parker, 1995, p. 209).

Despite being one key concept to understand cooperation (Henrich & Boyd, 2001), the study of functional punishment did not attract researchers' attention until recently (eg. Boyd, Gintis, Bowles, & Richerson, 2003; Sarah F Brosnan, Salwiczek, & Bshary, 2010; Clutton-Brock & Parker, 1995; Fehr & Gächter, 2000a; Joseph Henrich et al., 2006; Jensen, 2010; Marlowe et al., 2008). It is not until the last decade that it has been successfully explored in small-scale human societies (Hill, Barton, & Hurtado, 2009; Marlowe et al., 2008; Wiessner, 2005) and in different ethnic groups where "higher levels of punishment were significantly associated with higher levels of cooperation" (Henrich et al., 2006). With respect to animal societies, it has been demonstrated that functional punishment exists at least in second-party one-shot conditions (Clutton-Brock & Parker, 1995). For example, higher ranking individuals of rhesus macaques (*Macaca mulatta*) attacked lower ranking individuals when they did not give food calls (Hauser, 1992). However, it is not that clear whether non-human animals are able to

[16] A more subtle differentiation between reciprocal altruism and strong reciprocity is provided by Fehr and Fischbacher (2003, p. 785): "strong reciprocators bear the cost of rewarding or punishing even if they gain no individual economic benefit whatsoever from their acts. In contrast, reciprocal altruists, as they have been defined in the biological literature, reward and punish only if this is in their long-term self-interest. Strong reciprocity thus constitutes a powerful incentive for cooperation even in non-repeated interactions and when reputation gains are absent, because strong reciprocators will reward those who cooperate and punish those who defect." For our purposes, we will be more interested in reciprocal altruism, but without meaning that our subjects were intentionally calculating and foreseeing the consequences of the punishments they were applying.

[17] It is not necessary that individuals performing either second- or third-party functional punishment do it on purpose since the application of functional punishment can be learnt through operant conditioning.

form consistent reciprocal systems through iterated second-party punishment (i.e. defeating after defection).

One experiment evaluated punishment in a reciprocal interaction in chimpanzees through different conditions (Jensen, Call, & Tomasello, 2007b). The experiment presented food loss for the subject in three conditions. In the loss condition, the subject lost the food because the experimenter moved it away from him to an adjacent room with no partner. The unfairness condition was the same except for this time there was a conspecific that will enjoy the food in the adjacent room. Finally, in the theft condition the conspecific stole the food from the subject by pulling a rope. Importantly, the subjects could never recover the food but they could collapse the table where the food was. Collapsing the table made the food unavailable. If chimpanzees used functional punishment, they should pull from the rope more in the theft condition to punish the partner's non-cooperative action for the sake of future cooperation. Indeed, chimpanzees collapsed the table significantly more often in theft condition than in the remaining conditions. However, as the first author noted, "theft increased over time while retaliation decreased, suggesting that in the absence of immediate pay-offs functional punishment failed to enforce cooperative behavior" (Jensen, 2010, p. 2638)[18]. Importantly, each chimpanzee played 110 times (half as the subject), but only twice with the same partner in the theft condition (one as subject, another as thief). The authors measured functional punishment in between subjects, that is to say, the authors saw whether the subjects keep on retaliating after stealing despite changing partners. If the interaction between the same partners in theft condition only lasted two trials, new retaliations towards new partners would be better interpreted as altruistic punishment (the subject would not benefit from punishing the partner since they both will not interact again). As we have already mentioned that altruistic punishment is not common in animals, it would have been more interesting to see whether, in a more repetitive interaction in the theft condition the functional punishment persisted and eventually changed the partner's stealing behavior. One of the aims of the present dissertation is

[18] In an observational study, de Waal and Luttrell (1988) reported that chimpanzees, but not macaques, maintained a functional punishment reciprocal system. They also suggested that species-specific factors, such as hierarchy, could have an influence over the development of these systems: "harmful interventions were… reciprocal among chimpanzees only. This species showed a "revenge system", that is, if A often intervened against B, B did the same to A. In contrast, both macaque species showed significantly inversed reciprocity in their harmful interventions: if A often intervened against B, B rarely intervened against A. Further analysis indicates that the strict hierarchy of macaques prevents them from achieving complete reciprocity." (no page is reported because this literal citation comes from their abstract)

precisely to deeper exploring this through a different methodology. We studied the phylogeny of iterated functional punishment to achieve cooperation testing chimpanzees (*Pan troglodytes*) in an iterated ultimatum game.

To understand why we choose the ultimatum game, it is necessary to understand the underlying psychological mechanisms of functional punishment and cooperation. One of these mechanisms is inequity aversion (IA). Cooperation in resource distributional contexts means to balance outcomes or at least to not generate big differences between the participants' payoff. The IA is especially relevant here because it is the capacity to compare one's outcomes with those of the partner and reject the distributions that do not yield to similar payoffs (Brosnan, 2011; Brosnan, 2012). There are two types of IA: disadvantageous IA (the aversion to have less than the partner) and advantageous IA (the aversion to have more than the partner) (Brosnan & de Waal, 2014). In both cases, functional punishment would imply to incur a cost (i.e. loose some resource) when the partner has made unequal distributions as a way to protest against non-cooperation. The paradigmatic methodology to test IA and functional punishment is the ultimatum game.

In this game, a proposer decides how to split rewards with a partner, who as a responder in turn decides whether to accept or reject the offer. In the case of acceptance each player receives his or her corresponding split, but in the case of rejection nobody receives anything (Güth et al., 1982). Rejections are an example of functional punishment grounded on IA: the recipient observes the distribution, compares his outcomes with those of the proposer and, if he finds them very unequal, rejects. The rejection is costly for the responder as it is for the proposer, but the future benefit is that in consequent interactions the proposer may raise his offers to avoid new rejections. The mini-ultimatum game (MUG; Falk, Fehr, & Fischbacher, 2003) is a simplified version of the ultimatum game in which two splits are pre-assigned and the proposer needs to decide which one of the two he or she offers to the responder. The MUG has been used with children and chimpanzees because the proposers find it easier to split resources out of two distributions rather than out of a total amount of resources. Moreover, as the experimenter decides the preassigned distributions, the MUG also allows to including whatever comparison we may want to explore. For example, an 8/2 option (8 resources for the proposer, 2 for the recipient) may elicit disadvantageous IA, whereas a 6/9 may elicit advantageous IA. If the participants play through turn-taking in iterated interactions where proposer and responder roles are exchanged we can easily reproduce

the conditions for strong reciprocity to arise. Even though some previous MUG have already been conducted with chimpanzees (Jensen et al., 2007a; Kaiser et al., 2012; Proctor et al., 2013a), different methodological concerns precluded to have unanimous conclusions. The main purpose of Study 1 is overcoming these methodologies (see the corresponding Introduction) and play MUG with chimpanzees. We will also include the dictator game (DG), an economic game similar to the MUG with the exception of the recipient's role. Whereas in the MUG the recipient is active and can influence over the proposer's outcome through his acceptance/rejection, in the DG the recipient is passive and must accept whatever offer he receives. By playing iterated MUG and DG we should be able to see whether the absence of rejections in DG makes proposers being more selfish than in progressive trials in MUG. Thereby we will shed light upon the phylogeny of functional punishment in the formation of cooperative societies.

Revenge: the punishment without an apparent function

The kind of punishment I have discussed so far is one that produces some benefit. The underlying basic idea of punishment in Psychology was to think of punishment as motivated to promote certain behaviors while suppressing others, in other words, punishment was thought to be specifically addressed to deter (Butterfield, Trevino, & Ball, 1996). That was also the basis in the Law to apply punishments (eg. Kaufman, 2013; Tunick, 1992). But the function of punishment is not always that easy to determine. Take whatever example of those mentioned in the Introduction (eg. forty-seven rōnins, Medea, The Count of Monte Cristo, Kill Bill) and note that none seems to have been addressed to deter the offender or to promote cooperation.

Both Psychology and Law soon realized that people involved in conflicts were not mainly driven by the idea of future deterrence or by the reinforcement of cooperative rules (actually, deterrence through punishment is not even proven to be consistently effective (Honderich, 1969)[19])[20]. Indeed, deterrence means applying punishment for a future benefit, under prospective purposes (Hoffman & Spitzer, 1985), when it seems

[19] Some authors even propose that punishment can generate an endless cycle of violence. For example, "a vengeful act may provoke further aggressive reaction" (Stuckless & Goranson, 1992, p. 26).

[20] It is worth to say that although this is true at the proximate level (i.e. individuals might not be thinking about future deterrence) it may still be true at the ultimate level (i.e., natural selection has selected individuals to punish because those punished are less likely to repeat the offense at the population level). Besides, humans may actually be motivated by future deterrence but this would be a subconscious process that they would not be aware of, and they may not be able to report it if asked. However, I want to highlight the idea that the immediate function of functional spite is not obvious, contrary to functional punishment.

that many people are retributive, namely they apply punishment under retrospective purposes (Carlsmith, Darley, & Robinson, 2002). The discussion about the different functions of punishment according to prospective and retrospective motivations divided the legal theory into two different perspectives. On the one hand, the "utilitarian" school did not consider punishment as end in itself, but as an instrument addressed to deterrence (i.e. prospective). On the other hand the "just deserts" perspective did consider punishment as an end in itself addressed to make the offender suffer for the wrongdoing he/she committed (i.e. retrospective) (Carlsmith, 2006, pp. 437–438). Therefore, punishment is not only conceived for prevention: punishment can fail to produce any distinguishable benefits neither to the individual nor to the group. In those cases, it is denominated "functional spite".

Whereas functional punishment is synonym of retaliation (costly act that changes the offender's behavior), functional spite is synonym of revenge (costly act[21] that does not pursue the modification of the offender's behavior). Despite the large legacy of vindictive events both in the real and in the fictitious world that I have mentioned at the beginning of this Introduction; despite centuries of History, the study of revenge is devoid of much research. The studies conducted are surprisingly and unfairly limited (Frijda, 1994; Vidmar, 2000). As the eminent psychologist of emotions Nico Frijda stated:

> "No major psychological study has appeared on the topic during the last 70 or 80 years. A literature search from 1967 to 1991 yielded not a single study having "vengeance" or "revenge" as its main subject. (…) It should be a task of the psychology of emotion to devote attention to the properties of wrath, that is, of such anger as leads to vengeful fantasies and actions" (Frijda, 1994, pp. 264–265).

The interesting fact is that "people take revenge even if it is costly for them and yields no future material reward" (Fehr & Gächter, 2000, p. 159). The action of revenge does not seem to be useful with regard to the object of loss. Thinking of a particular case of revenge, it is unclear why Hamlet killed Claudio. Killing Claudio did not heal any of

[21] Cost can be considered of different modalities (i.e. time, money, physical effort (Cota-McKinley, Woody, & Bell, 2001; Crombag et al., 2003; Elster, 1990) or risk (i.e. being condemned). Whichever the modality, it is essential that revenge is costly for the avenger "to be classified unequivocally as revenge" (Schumann & Ross, 2010, p. 119).

Hamlet's father lethal wounds and therefore made nothing for his recovering. It did not have a deterrent function either since Claudio was not alive to change his future behavior. The inability to replace a previous loss plus the cost that revenge entails (namely, "its apparent uselessness and occasional self-destructiveness" (Frijda, 1994, p. 265)) but the urge to be done anyway have led many authors to use the expression "the paradox of revenge" or to describe it as "irrational" (Crombag, Rassin, & Horselenberg, 2003). This reference to the purpose of revenge is not pointless. Revenge is a striking and contentious evolutionary fact. It opens to question why we are willing to experience functional spite, which is similar to asking about what makes revenge so beneficial to ourselves that it has been able to persist over time (and maybe even extend to other species). Precisely the wide distribution of revenge may have hindered its study "because everyone has instantly available intuitions that supply their own definitive answers about this, people generally have failed to notice that there are no research-based answers to this rather fundamental question about human nature" (Carlsmith et al., 2002, p. 285). A quick look to some of the definitions provided by different authors across time might help to find the functions of functional spite (i.e. revenge) (highlighted in bold):

> "Vengeance is an act designed to harm someone else, or some social group in response to a feeling that oneself has been harmed by that person or group, whereby the act of harming that person or group is not designed to repair the harm, to stop it from occurring or continuing in the immediate confrontation, or to produce material gain. (…) [Despite] its apparent uselessness and even occasional self-destructiveness (…) the immediate purpose of revenge, as the definition states, is to **make the object of vengeance suffer**" (Frijda, 1994, pp. 265-266)
>
> "Unlike justice, the desire for vengeance-driven punishment is personal. The act of revenge has the essential purpose of **giving the avenger relief** from a feeling of discomfort" (Ho, ForsterLee, ForsterLee, & Crofts, 2002, pp. 366–367).
>
> "When faced with a prototypical wrongdoing action, a harm intentionally inflicted on another by a perpetrator, people assign punishment to **give the perpetrator his or her just deserts** rather than to achieve any future utility" (Carlsmith et al., 2002, p. 295)"Whereas revenge is motivated by a yearning to

> see a transgressor suffer, punishment is motivated by a desire to improve a transgressor's future behavior. (…) We emphasize the intention **to see the transgressor suffer** in our conceptualization of revenge" (Schumann & Ross, 2010, p. 1194)
>
> "Is the goal to reform subsequent behaviour of the target? Alternatively, is the goal more abstract, such as achieving cooperative norms? Perhaps the motives behind punitive acts are antisocial, having **the suffering of the target as the primary goal** with any positive effects being unintended by-products" (Jensen, 2010, p. 2636)

Most of these definitions place the suffering of the perpetrator as the main purpose of revenge. Therefore functional spite might preserve the adjective "functional" because, in Jensen's words:

> "What distinguishes functional spite from functional punishment is that functional spite does not require any change in the target's subsequent behaviour. The end goal is the harm incurred by the target. There may be indirect benefits—otherwise the behavior would not be functional—but these are less tangible than for functional punishment. Whereas functional punishment emphasizes the delayed benefits to the punisher, functional spite emphasizes the immediate costs to the target; negative consequences for the target are the *raisons d'être* for spiteful acts. Functional punishment is a means to an end; functional spite is an end in itself. The benefits that accrue to the actor would therefore be indirect; the target's loss is the actor's gain." (Jensen, 2010, pp. 2642–2643)

One potential explanation for the selection of functional spite is that it works as a form of altruism (Gardner & West, 2006; West & Gardner, 2010). In other words, reducing the offender's fitness benefits both related and non-related individuals. Some authors propose that the existence of functional spite allows to generate an "hyper-competitiveness" (eg. Hill et al., 2009; Jensen, 2010) that is so uniquely human that Solomon described it as an "undeniable aspect of the way we react to the world, not as an instinct but as such a basic part of our worldview" (Solomon, 1994, p. 308). To check whether that is the case, another aim of the present dissertation was to test children through a one-shot MUG. We chose the MUG because it is considered "the

most vivid game to demonstrate negatively reciprocal behavior" (Fehr & Gächter, 2000b, p. 161). Few studies have conducted the MUG with children (Gummerum & Chu, 2014; Harbaugh et al., 2003; Leman, Keller, Takezawa, & Gummerum, 2009; Sutter, 2007; Wittig et al., 2013), and all of them had serious methodological shortcomings (for more information, see Study 2) that produced confounded results. As a result, the ontogeny of the underlying mechanisms of revenge remains unclear. With this one-shot MUG we could test revenge because there were no future interactions with the partner and therefore rejections could not serve as deterrence (Fehr & Fischbacher, 2003), but should be triggered by functional spite. This is precisely the difference between this MUG and the iterated MUG that we previously described for chimpanzees. It is true, though, that it seems to exist more motives to reject in MUG than "making the partner suffering". I will discuss them straightaway, but first I will highlight the conditions we used because their configuration could guide the different motives to reject.

Importantly, the conditions we used in the MUG with children were the conditions from the original MUG study (Falk et al., 2003) in which not only outcomes, but also intentions, were taken into account by responders before rejection. The original conditions had a default option (8/2) pitted against different conditions that varied in the level of IA they generated. For example, in the condition "No choice", the proposer could decide what to offer between two identical options (i.e. 8/2), whereas in the condition "Fair", the proposer could decide between 8/2 and 5/5. If only outcomes were important, the 8/2 option should be equally accepted/rejected across all conditions. However, if outcomes plus intentions matter, in No Choice condition, despite 8/2 generated a disadvantageous IA for the responder, he should see that the proposer did not have any purposeful intention to produce the unequal payoff (both options were identical) and should be more willing to accept. By contrast, in the Fair condition, an 8/2 offer would generate an intentional disadvantageous IA because the proposer could have offered the better payoff that was available and more rejections should arise. Indeed, the authors showed that intentional offenses were more likely to be reciprocated

than non-intentional[22]. The combination intentions/outcomes is very relevant in the study of revenge[23]; this is the reason why we opted for a one-shot MUG with children.

Coming back to the previous question, what are the motives to reject in a one-shot MUG apart from making the partner suffering? One motive is the enforcement of social norms. Social norms may estipulate how the sharing should be between two members of the same society according to the common idea of fairness; therefore the rejection could be the reaction to the unfulfillment of that norm (Raihani & Mcauliffe, 2017). This kind of rejection implies the existence of a social norm; the ability to detect when it has been broken and the willingness to protest for not having being applied. We know that from very young children are able to protest when social norms are transgressed and that they can even act out of spite when others get more than them (Blake & McAuliffe, 2011; McAuliffe, Blake, & Warneken, 2014; Schmidt, Rakoczy, & Tomasello, 2012; Sheskin, Bloom, & Wynn, 2014). Another reason to reject is IA. Rejections based on IA imply comparing the recipient's and the proposer's outcome. As a result of that comparison, the recipient does not like to receive less or more than the partner and prefers to avoid that difference by rejecting (zero for both). Therefore, IA compares the participants' outcomes, regardless of social abstract ideas. A different reason to reject is frustration. Rejections based on frustration imply comparing two different quantities and judge that one is lower than the other. As a result of this comparison, the recipient does not like to receive less than what was further available and prefers to avoid conformity. Therefore, frustration compares available outcomes, regardless of what the partner has and regardless of social norms. Frustration can be mixed with the inclusion of intentionality. If the recipient knows that the donor had the power to decide which option to offer and he purposefully decided to choose the lesser amount, then we will not know whether the

[22] This is also true for children as young as 3 years of age, who intervene when punishments are applied to accidental rather than intentional actions (Chernyak & Sobel, 2016).

[23] Research has shown how intentionality also matters in exchange games. For example, in a resource distribution game, subjects behaved differently depending on whether they received an unequal outcome from a computer-roulette or from another human partner (Blount, 1995). Computers are not intentional agents, therefore their decisions in terms of how to distribute resources are not considered unfair and people accept them easily (i.e. do not show revenge). However, human partners have the capacity to decide whether doing something or not, therefore their actions are considered differently and can arouse negative reciprocity. Similarly, in another experiment (Keysar, Converse, Wang, & Epley, 2008) two groups of subjects played a reciprocation game after receiving 50$ from an anonymous player. The first group was told that the money came from an act of giving, namely the anonymous partner had 100$ and decided to give them 50$. The second group learnt that the money came from an act of taking, namely both subjects were assigned 100$ and the anonymous partner decided to take 50$ for him. Even though the outcome earned was identical for both groups, the individuals in the group of "giving" reciprocated more money than the individuals in the group of "taking".

rejection was motivated by the comparison between quantities (frustration) or by the desire to punish the donor for his low generosity (intention).

Despite the variety of the reasons to reject, all of them have the loss of the responder and the partner's outcomes in common. This means that the responder is incurring a cost to cause a cost, when in principle a rational maximization interest would advise against rejection (Persky, 1995). That is why we consider that this one-shot MUG with conditions that combine intentions and outcomes can still shed light upon the ontogeny of functional spite (taking into account the diversity of motives to be applied).

Schadenfreude: the enjoyment of punishment

Similar to functional spite, it is difficult to see what the function of Schadenfreude is. Recalling the beginning of this Introduction, attending to watch the execution of your beloved's murderer could trigger pleasure, but this pleasure will not help to recover your beloved. However, it seems that the urge to attend the execution is imperative. The media coverage and the exhaustive research interviews conducted by Madeira (2012) to the attendees of the real execution are good evidence:

> "Participants were adamant… that closure could never occur because what was lost could never be regained… Oneta Johnson, 34, said… watching Mr. McVeigh die… might make things better… Raymond Washburn, 54, came to the viewing even though he is blind."[24]

Note how Oneta foresaw some delight in the consequences of the punishment and how the blind man incurred costs (i.e. trip, time, money) to attend the execution despite his inability to watch it. These two cases reinforce the idea of humans' need for seeing negative acts reciprocated as well as humans' joyful experience when perpetrators receive punishment. Some previous research in criminology showed that people reaction to punishment is more emotional than rational-based (eg. Ellsworth & Gross, 1994). This is also evidenced by Neurosociology: the experience of emotions is a likely proximate cause that sustains cooperation and motivates costly punishment of antisocial others in humans (de Quervain et al., 2004; Fehr & Gächter, 2002; Singer et al., 2006). Whereas seeing others suffer typically induces emotional states such as empathy (Singer

[24] Yarley, J. (2001, June 12). "The McVeigh Execution" [Report]. Retrieved from http://www.nytimes.com/2001/06/12/us/the-mcveigh-execution-oklahoma-city-execution-on-tv-brings-little-solace.html

et al., 2004), which in turn is a powerful motivator for altruistic helping (Batson, 1991; Batson, Duncan, Ackerman, Buckley, & Birch, 1981; Hein, Silani, Preuschoff, Batson, & Singer, 2010), this can be radically undermined and change to feelings of pleasure when the suffering victim was previously antisocial or perceived as an outgroup member (Hein et al., 2010; Singer et al., 2006).

For example, in one recent study (Cikara & Fiske, 2013) the authors demonstrated that the supporters of two baseball teams traditionally confronted (Red Sox and Yankees) exhibited neural correlates of pleasure (i.e. ventral striatum activation) while watching how the competitive rival failed. Therefore the Red Sox's supporters objectively (neural correlates) and subjectively (self-reports) felt pain when the rival scored (Cikara & Fiske, 2013) but that pain was reciprocated when their team scored back (revenge), what triggered feelings of pleasure (Schadenfreude) in the supporters[25]. Such signals of reward have been shown to be critical predictors of a subsequent absence of helping and desire for revenge and punishment (Hein et al., 2010; Singer et al., 2006). Therefore, it seems that revenge is the behavioral response to antisocial perpetrators whereas Schadenfreude is the emotional response. With this assertion I am not implicitly saying that first is revenge and then, as a consequence comes Schadenfreude. At this respect, as I have just mentioned, the literature is ambiguous. It can also be that Schadenfreude is the motivation to take further revenge.

Importantly, here I am excluding "Schadenfreude" as the enjoyment of whoever's misfortune that I also mentioned in the Introduction (eg. TV program Fūun! Takeshi Jō). This is because I am going to focus on the enjoyment of revenge and for revenge to exist a previous offender is needed. However, it is worth mentioning that the enjoyment of unknown's and offender's misfortune are preceded by the activation of different parts in the brain. This might be a sign for different evolutionary functions. Funny unknown's misfortune triggers the areas in charge of detecting incongruence (Samson, Hempelmann, Huber, & Zysset, 2009) and the ulterior reward feeling activates areas of the brain different from the rewarding areas activated while watching pleasant images (Neely, Walter, Black, & Reiss, 2012). Neely and colleagues used videos of children being catapulted into the air from an inflatable couch as funny unknown's misfortune situation (i.e. children jumping should not fall but if they do the spectator detects the

[25] For an interesting discussion on the concept of Schadenfreude applied to sports, please refer to "Chapter 8: Schadenfreude in sports: envy, justice and self-esteem" (McNamee, 2008, pp. 145-161).

incongruence and laughs) and concluded that "the temporo-occipito-parietal junction (TOPJ) activation may be specific to humor processing and not part of a general constellation of neural activity in response to reward". These findings led to the "benign violation theory" which proposes that humor occurs when and only when three conditions are satisfied: (1) a situation is a violation, (2) the situation is benign, and (3) both perceptions occur simultaneously (Warren & Mcgraw, 2015). However, this theory cannot account for offender's misfortune because the spectators in that situation are willing to see a non-benign situation (i.e. punishment) enacted. In fact, the enjoyment that they feel comes from finding congruence rather than incongruence between a previous event and the punishment enacted. Besides, the ulterior reward feeling does activate the traditional rewarding system (Singer et al., 2006). Therefore, Schadenfreude is the enjoyment of others' misfortune but it activates the brain differently depending on whether the affected agent was unknown (incongruence) or a previous offender (congruence). Evolutionarily, it makes sense that both events are rewarding and it might be indicative of different functions: unknown's misfortune might be useful for reinforcing the ability to understand physical and social rules whereas offender's misfortune might be useful for reinforcing defensive actions (i.e. revenge) against those who can harm the individual.

Indeed, some authors do have a strict conception of Schadenfreude that applies only to certain particular scenarios. Cikara and Fiske recalled in their study (2013) that Schadenfreude is predicted in different conditions (Smith, Powell, Combs, & Schurtz, 2009): (1) when the misfortune benefits the observers (Smith, Eyre, Powell, & Kim, 2006); (2) when the misfortune seems deserved (Feather & Nairn, 2005; Feather, 2006); and (3) when the misfortune befalls an envied target[26] (Schoek, 1987; Smith et al., 1996; Takahashi et al., 2009). For the purpose of the present dissertation I will focus on (2), since the deservingness is the closest scenario to revenge (Frijda, 1994). Moreover, I am going to consider Schadenfreude as the consequence of seeing revenge enacted. This means that in our experiment subjects will not directly execute their vengeance. A

[26] A funny example of Schadenfreude coming from envy appeared in one episode of The Simpsons' cartoons (3x03, "When Flanders Failed"). Part of the script is the following:
Homer (H): Ha, ha, ha. I'm telling you: Flander's store was deserted! (…)
Lisa (L): Dad, do you know what Schadenfreude is? (…) It's a German term for shameful joy: taking pleasure in the suffering of others.
H: Oh, c'mon, Lisa! I'm just glad to see him fall flat on his butt. He's usually all happy and comfortable and surrounded by loved ones.
The scene can be watched here: https://www.youtube.com/watch?v=B01e7n4RzZc (Last accessed: 2016 January 2).

third-party will punish the offender and the subjects will be witness of that punishment. We will be interested in coding the reactions that the subjects have during the punishment. On the other hand, the popular quote "Revenge is always sweet" does not seem to always apply. The evidence of the emotional consequences of revenge are mixed: while some authors have proved the delights derived from revenge (de Quervain et al., 2004; Funk, McGeer, & Gollwitzer, 2014; Gollwitzer, Meder, & Schmitt, 2011), others have shown that avengers also experience negative emotions (Carlsmith, Wilson, & Gilbert, 2008; Lambert et al., 2014) (aroused, for example, by the memory of the event, namely recalling the moment when the harm was produced). The current perspective, therefore, is the so-called "bittersweet model" of revenge (Eadeh, Peak, & Lambert, 2017), which is grounded on the need for exploring the emotional experience of revenge in a multi-modal way. For the purpose of this dissertation, I will code both negative and positive reactions as well as I will include both emotional and behavioral measures in the corresponding experiments.

Even though there are some studies measuring Schadenfreude, they have generally used videos and rated the feelings or coded the neural activity of the subjects (Cikara & Fiske, 2013; Smith et al., 1996; Steinbeis & Singer, 2013; Takahashi et al., 2009). Therefore realistic costly situations in which the subjects had to pay some cost to enjoy seeing others suffer are absent with children and chimpanzees populations, thus making the ontogeny and phylogeny of the enjoyment of punishment unclear. In particular, we ran two experiments. In both of them we presented a prosocial and antisocial agent. We wanted to see whether the punishment of the prosocial agent arises reactions compatible with empathy and whether the punishment of the antisocial provoked reactions compatible with Schadenfreude. Following a multimodal approach, we coded whether individuals incurred a cost to escape from the prosocial punishment or incurred a cost to continue watching the antisocial punishment (physical action) and whether they frowned and smiled (bittersweet model of revenge in the case of children) or screamed and were piloerected (in the case of chimpanzees).

The difference between the two experiments was the level of the subjects' involvement with the agents. In Study 3 both agents directly interacted with the subjects. Therefore the ulterior punishment could be considered as an indirect revenge (in the case of the antisocial agent) where a third person, the punisher, executed the punishment "on behalf" the subject. In Study 4, however, the subject was a bystander. Subjects saw how

the agents interacted with a conspecific and then, again, the punisher executed the punishment to both of them. With regard to the level of involvement, therefore, Study 3 was a second-party situation whereas Study 4 was a third-party scenario.

We run Study 4 only with chimpanzees, not with children. This is because we already know that humans intervene in third-party transgressions, namely "people punish norm violators not for what they did to the punisher but for what they did to others" (Fehr & Fischbacher, 2003). For example, humans incur cost to punish non-cooperators in public good games (Fehr & Gächter, 2000a) and in iterated prisoners' dilemmas (Fehr & Fischbacher, 2004) even when they never suffered the consequences of the offender's cheating. This is also true for children, who also intervene in third-party moral transgressions (Jordan, McAuliffe, & Warneken, 2014; Vaish, Missana, & Tomasello, 2011 but see Gummerum & Chu, 2014). However, this third-party intervention is very rare in non-human primates. Even though it is demonstrated that some non-human primates recognize non-cooperators from a third-party perspective (Anderson, Kuroshima, et al., 2013; Anderson, Takimoto, Kuroshima, & Fujita, 2013; Kawai et al., 2014; Russell, Call, & Dunbar, 2008), it is not consistently demonstrated that they intervene in the form of third-party punishment. For example, chimpanzees did not incur any cost to punish those who stole food from others in an experimental set-up (Riedl et al., 2012). It is true that there are some cases of policing (i.e. costly third-party mediation in fights) between non-human primates in natural environments (Flack, Girvan, de Waal, & Krakauer, 2006; von Rohr et al., 2012), but they are very rare. We only run Study 4 with chimpanzees because we wanted to explore the phylogeny of third-party (indirect) punishment. Contrary to humans, chimpanzees might not be very willing to intervene when they need to perform the third-party punishment (response level), but we do not know whether at least they have the underlying motivation to see third-party punishment enacted (motivational level).

Our aim with the one-short third-party Schadenfreude experiment was to fill this gap and see whether chimpanzees were motivated to (at least) costly witness third-party antisocial punishment. Based on the low frequency of third-party interventions our hypothesis was that chimpanzees would not be willing to incur costs in such punishment. But in case subjects were willing to incur cost to watch the offender of a conspecific being punished the interpretation of the results would be challenging. The interpretation would not be compatible with feeling pleasure to see the offender

suffering because the harm was not directly received. The interpretation would be compatible with the satisfaction of seeing norms being applied (Raihani & Mcauliffe, 2017). Nevertheless, in case chimpanzees incurred some cost, we would be very cautious to interpret the result as synonymous of the existence of social norms in non-human primates, since this has not been demonstrated in any other species different from humans yet (Brosnan, 2012).

Next, we provide a summary (Table 1) of the studies that comprised the present dissertation with a brief explanation of their objectives and methodology.

Our studies

The aim of the present dissertation was to study punishment from different perspectives across four different experiments. This means that in all of them there was an offender and there was a subject. The offender was the individual who provoked some harm (not always to the subject, as it will be explained below), whereas the subject was the individual who could perform actions as a response to that harm. Importantly, the subject always had to incur some cost to perform those actions (see fourth column in Table 1).

As it can be seen in Table 1, in two of the experiments the subjects' actions were directly addressed to the offenders (UG: rejection), whereas in the remaining two the subject could just participate as a bystander of the punishment inflicted to the offender by a third-party (Schadenfreude). This division between direct and indirect experiments is relevant because with direct experiments we focus on responses, namely the subject's willingness to get even/retaliate against the offender, whereas with indirect experiments we focus more on motivation, namely the subject's willingness to watch the enactment of revenge.

The MUG experiment allows for measuring responses because the recipients can execute some action (i.e. reject) towards the responsible of the harm, namely, the proposer who offered unequal distributions of resources. In those experiments we analyzed both the proposer and the responder's responses, however for the purpose of the present dissertation I am going to focus only on the responder's responses. In particular, we were interested in seeing whether rejections were triggered by the previous interactions with the partner (in the case chimpanzees' iterated MUG), by IA

Table 1.
Summary of the studies of the present dissertation and how we studied punishment in each one.

Study	Did the offender harm the subject?	Could the subject directly punish the offender?	Did the subject incur some cost?	Was their interaction (subject/offender) repeated?	Denomination	What does the subject obtain from the offender's punishment?
Study 1 **UG** **apes**	Yes, Second-party Proposer offered few/no food	Yes, Direct Rejection	Yes Lost food	Yes Iterated	Retaliation Functional punishment	Change Behavior Promote cooperation
Study 2 **UG** **Children**	Yes, Second-party Proposer offered few/no stickers	Yes, Direct Rejection	Yes Lost stickers	No One-shot	Revenge Functional spite	Pleasure Norms enforced?
Study 3 **Schadenfreude** **children / apes**	Yes, Second-party Antisocial took food away	No, Indirect Subject could witness how third-party punished the antisocial agent	Yes Lost coins / Physical effort (open door)	No One-shot	Indirect revenge Schadenfreude	Pleasure Norms enforced?
Study 4 **Schadenfreude** **apes**	No, Third-party Antisocial took food away from a stooge	No, Indirect Subject could witness how third-party punished the antisocial agent	Yes Physical effort (open door)	No One-shot	Indirect altruism? Schadenfreude	Norms enforced?

(disadvantageous or advantageous) or by frustration (in the case of both MUG). They could also reject for the unfulfillment of social norms, but as this was difficult to disentangle, we leave a question mark in the corresponding cell of Table 1. To explore IA and frustration we had different conditions in which we varied the quantities of the distributions that the proposers could offer. A more detailed explanation of these conditions can be found in the respective Conditions sections.

The Schadenfreude experiments allow for measuring motivation because the subjects could not execute actions towards the responsible of the harm but they could incur some costs to witness how the offender is being punished by a third-party. Our aim was to know whether children and chimpanzees had a preference for watching antisocial as opposed to prosocial agents being punished. In particular, we were interested in exploring whether empathy (when punishing prosocial agents) and Schadenfreude (when punishing antisocial agents) emerged, since both "other's fortune emotions" (Ortony, Clore, & Collins, 1988) have an influence over the motivation to execute/to enjoy punishment (Jensen, 2010).

Another differentiation between our four experiments is whether the subject received the harm directly from the offender or not. The level of directedness of the harm is relevant for the execution/enjoyment of punishment. Recall that bystanders of the Oklahoma bomb attack did not receive any direct harm from the explosion but were equally willing to watch McVeigh's execution. In three of our experiments the offender directly harmed the subject (second-party punishment), whereas in the remaining one the offender harmed a conspecific (third-party punishment). In this sense it was also important whether the interaction between offender and subject was repeated. We ran one-shot second punishment (MUG with children); iterated second-punishment (MUG with apes); one-shot second- and third-punishments (Schadenfreude with children and apes). This division between second and third-party punishment with the addenda of one-shot or iterated is relevant because the function of the punishment varies. Second- or third-party punishments are usually aimed to promote cooperation if iterated (functional punishment). The function is more ambiguous if one-shot, being perhaps the satisfaction of seeing the offender suffer (especially in second-party conditions) or a potential individual willingness to seeing some social norm being endorsed (i.e. offenders must always be punished) (especially in third-party conditions) (functional spite). Concretely, in the iterated MUG we wanted to see whether chimpanzees can

endorse cooperation through functional punishment whereas in the one-shot MUG and in the one-shot second-party Schadenfreude we wanted to see whether functional spite was present. Again, in the latter case, as it is difficult to infer just by looking at the rate of rejections or the costly opening of the doors whether the function of the functional spite was to payback or to endorse social norms. Therefore, we prefer to leave a question mark in the corresponding cells of Table 1.

Study 1:

MUG and DG in chimpanzees

State of the art

Current interest in the evolutionary roots of fairness and its psychological underpinnings have led researchers to confront pairs of individuals of various species (mainly primates) with various social dilemmas (for reviews, see Jensen, 2010; Silk, 2009) including several versions of the UG (Jensen et al., 2007a; Kaiser et al., 2012; Proctor et al., 2013a). Following the mini-ultimatum procedure (Falk et al., 2003), Jensen and colleagues (2007a) presented pairs of chimpanzees with preselected pairs of quantities (e.g., 5/5 vs. 8/2). The proposer could select one of the pairs by pulling a rod that brought the offer halfway. Then, the recipient could accept it by pulling another rod that delivered the offer to both subjects or reject it by not pulling during the next 60s, thus ending the trial without any food within reach. Kaiser and colleagues (2012) tested another group of chimpanzees and bonobos using a similar methodology except that in some conditions the proposers could also "steal" some of the food originally allocated to the recipient before making an offer to the recipient, to see whether this enhanced rejections.

In both studies, proposers did not incur cost to make equal offers whereas recipients showed no IA since they never rejected non-zero outcomes. Consequently, unlike humans, chimpanzees and bonobos behaved as rational maximizers. With regard to recipients, one argument against this conclusion was the fact that 0-options were accepted approximately half of the time in Jensen et al's study (2007a). According to some authors, chimpanzees might not have been rational maximizers if they had accepted 0 as outcome (Brosnan, 2013) but they had not understood the procedure. According to others (Henrich & Silk, 2013), rejecting 0-option half of the time implies behaving at chance, which is compatible with rational maximizing as both accepting and rejecting leads to zero outcome. A different explanation came some years later from Smith and Silberberg (2010). They found that apes' data was reproducible in humans if increasing the delay to reject from 1 to 5 minutes. When humans were forced to wait 5 minutes to reject they tended to accept whatever payoff to make the next trial start and therefore increase their likelihood to obtain something. After all 60s may have been too long to wait for chimpanzees in Jensen et al (2007a), and the acceptances of 0-options reinforced the idea of rational maximizers as the chimpanzees might have rejected looking for a new opportunity to obtain something. Anyway, if chimpanzee responders did not reject, they were not providing signs of IA or functional spite. Were

chimpanzees indifferent to differential outcomes and therefore reluctant to undertake actions against inequalities?

Some years later Proctor et al. (2013a) compared chimpanzees' responses in an MUG and in a preference test[27] using a token exchange paradigm. They wanted to see whether the active/passive role of the conspecific could affect the proposer's offers. To do so, chimpanzees were trained on the value of two tokens, each allocating a different amount of food to the proposer and the responder (5/1 vs. 3/3).

In their MUG, proposers selected one out of the two tokens, gave it to the responder who could then either return it to a begging experimenter (accept) or keep the token during the next 30s (reject). In their preference test, proposers gave tokens directly to the experimenter while a naïve passive recipient sat in the adjacent cage. Despite the methodological shortcomings argued against this preference test (Henrich & Silk, 2013), they found that proposers selected the 3/3 token more often in the MUG than in the preference test. The interpretation for this result was controversial. Henrich and Silk (2013) pointed out that the change towards 3/3 was not different from chance in two out of the three dyads and only matched very culturally restricted UG results in humans (those found in Western human undergraduates). In response, Brosnan and de Waal (2014) claimed that this change of behavior between conditions reflected second-order IA as chimpanzees might have anticipated a conflict (i.e. rejection) and have prevented it by offering more generously.

The authors' interpretation is interesting for our purposes because it is related to the underlying mechanisms of functional punishment, which is IA. Whereas we had defined IA so far as the comparison that the subject does between his and those of his partner's, second-order IA is a step further. Second-order IA consists of "recognizing when they [the proposer] receive more, as this allows them to forestall first-order IA reactions in their partners and thereby maintain a successful cooperative relationship". In other words, second-order IA is the capacity to foresee that the partner is going to reject future unequal offers driven by his disadvantageous IA. If this second-order IA existed,

[27] The authors presented the preference test as a dictator game (DG). The difference between the DG and the MUG is the role of the responder. In the DG, the responder should accept whatever offer from the proposer. A rational maximizer should offer zero in DG whereas he should strategically offer the least possible amount in the MUG to avoid partner's rejections.

chimpanzees would be behaving as cooperative partners, which is the purpose of functional punishment after iterated interactions.

However the existence of second-order IA is unclear since rejections never occurred during their experiment. In their own words: "The change in choices was apparently spontaneous, occurring without any refusals by the partner" (Proctor et al., 2013a, p. 2071). Hence, there is no experimental evidence for that potential anticipation. Besides, the authors themselves recognized that subjects may have not refused because "neither species was explicitly trained that refusal was an option" (Proctor et al., 2013a, p. 2072). Importantly, these authors are defending the emergence of cooperation not based in learning by partner's punishment (functional punishment) but by a sudden insight in the proposers. This is not a parsimonious explanation, and that is why we are reluctant to conclude that chimpanzees can form cooperative societies and that they do so without the need of functional punishment, as it has been largely demonstrated in humans.

Some of the methodological shortcomings of their experiment that provided more plausible explanations for the absence of rejections were the absence of 0-options or the inclusion of a human experimenter with a begging hand. On the one hand, as chimpanzees never reject offers above zero, facing them with a 0-option would have helped to explore the full scope of possibilities in which chimpanzees might react to extreme disparity by rejection. In case the proposers changed their future offers in accordance to those rejections, then second-order IA and cooperative interactions could be further explored. Indeed, that would have been a clear case of functional punishment (i.e. punishing the mean proposer changes his future behavior). On the other hand, captive chimpanzees are used to return objects when humans beg them with their hands (i.e., keepers requesting transfer of objects). Therefore, it is difficult to disentangle whether rejections were absent because the apes saw a begging hand or because they actually found the payoff tolerable. Future studies should manage without the experimenter as much as they could.

So far, the phylogeny of cooperation in bargaining contexts is still ambiguous. There is mixed evidence upon whether chimpanzee proposers can be considered consistently prosocial in the MUG as opposed to consistently selfish in situations where they do not depend on their partner's response (Jensen et al., 2007a; Kaiser et al., 2012; Proctor et al., 2013a). Moreover, there are different interpretations about the absence of rejections

in recipients (Jensen, Call, & Tomasello, 2013; Proctor, Williamson, de Waal, & Brosnan, 2013b). Finally, it is still uncertain whether second-order IA might arise in these contexts. Introducing iterated procedures might be very informative here. The goal of Study 1 is to shed light upon this issue and contribute to compare chimpanzee responses in iterated bargaining games (i.e. MUG and DG. Recall from the Introduction that both are distributional games between a proposer and a recipient, but the main difference is the role of the later. In the MUG the recipient can decide whether accepting or rejecting the proposer's offer whereas in the DG the recipient is passive). The MUG will provide us with information about the IA in responders, therefore the rate of acceptances and rejections will be the priority result. In case that rejections happened (both of zero and of non-zero outcomes since both lead to proposer's loss of food), then we should look at whether proposers actually changed their successive offers. Functional punishment would be at work in case they did. The comparison with the DG at this point is very relevant, because in the DG the responders cannot apply any punishment. Therefore, proposers should only change their behavior in MUG, but not in DG. In other words, in the results we should find an increased generosity in MUG.

In order to overcome all the methodological shortcomings that we have mentioned so far, we did the following (see Table 2 for a summary). Players faced each other and the proposer chose between two options that delivered food to each partner. We used an ABACA design that alternated between non-social (A) and social games (B and C represented DG or UG) –a feature that allowed us to obtain a reliable estimate of the baseline tendency to select each option in the absence of a partner to compare it with their responses in social games. Moreover, the inclusion of the non-social condition allowed us to assess whether subjects understood the game and displayed a stable preference throughout testing. Furthermore, we reduced the time for rejection to 15s and the intervention of the experimenter during the games. We introduced conditions to test for rejection, cost for the proposer and IA: a 0-option (6/0) to explore the occurrence of rejections in absence of any outcome, which could work as functional punishment; 5/x and 6/x options (cost/no cost) to explore whether proposers were more willing to incur costs in any of the social games[28]; x/3 and x/9 options (low/high gift) to explore whether

[28] Proposers should incur more cost in the MUG than in the DG. If that was the case, the absence of rejections would confirm that they played strategically. With these data we would have demonstrated that proposers anticipate future rejections, which was the conclusion of Brosnan and de Waal (2014) but without the shortcomings of their study (Proctor et al., 2013a): the support of a well-constructed DG, an

proposers took into account their partner's payoff in their offers and to explore whether recipients rejected based in advantageous (rejection of high gift) or disadvantageous (rejection of low gift) IA. This resulted in a default option (6/0) pitted against one of the remaining options (6/3, 5/3, 5/9, 6/9, see Table 4 for further information). Finally, chimpanzees played reciprocal trials (i.e. every dyad played the same condition switching roles) and we scored any communicative act in the meanwhile to see whether second-order inequity aversion or reciprocity occurred.

Table 2.
Methodological comparison between the MUG conducted so far. We exclude Kaiser et al study because the inclusion of the theft condition made their aim different from ours.

	Jensen et al, 2007	**Proctor et al, 2013**	**Our study 2016**
Model	ABA Training+Test+Postraining	B Preference Test+UG	ABACA Training+Test+Postraining
Non-social condition	Yes	No	Yes
Materials	Big apparatus	Tokens	Simple apparatus
Experimenter	No intervention	High intervention	Low intervention
Physical proximity	No (adjacent)	Yes (interaction)	No (but free to approach)
Quantities	(8/2) + (8/2, 5/5, 2/8, 10/0)	3/3 + 5/1	(6/0) + (6/3, 6/9, 5/3, 5/9)
Equity option	Yes (5/5)	Yes (3/3)	No
Zero option	Yes (10/0)	No	Yes (6/0)
Disadvantageous inequity option	Yes (2/8)	No	Yes (6/9, 5/9)
Advantageous inequity option	Yes (8/2, 10/0)	Yes (5/1)	Yes (6/0, 6/3, 5/3)
Type of rejection	Not pulling from a rod	Not returning token to a begging experimenter	Not pulling from a Velcro handle
Rejection delay	60 s	30 s	15 s
Kinship	Not controlled	Not controlled	Controlled
Cost comparison	No	No	Yes
Reciprocity (change of roles)	No	No	Yes
DG comparison	No	Yes (different methods)	Yes (same methods)

ABACA design, no intervention of the experimenter and less time to reject. In case that proposers incurred more cost in MUG than in DG after rejections, then we would have demonstrated that functional punishment was at work, because the inability to refuse in the DG kept proposers' offer immutable.

Study 1:
sample

Chimpanzees tested

Our sample is composed by six subjects (4 males; age range= 8 to 21 years) playing in dyads. The individuals of each dyad belonged to the same social group. In the zoo of Leipzig there are two different social groups. Group A is a multi-male group with one alpha male. It consists of 16 adults (nursery reared: 7; sterilized: 2) and 2 young individuals. This composition barely changed over the course of the present study (1 birth in 2015). Group B consists of 7 adults with one sterilized male.We ran 4 dyads controlled by kinship: Lobo+Lome and Kofi+Kara (step-brothers), Lobo+Kofi and Alex+Jahaga (unrelated). Two subjects (Lobo and Kofi) played twice because we wanted to explore whether they changed their behavior depending on the partner they were playing with[29]. Table 3 shows detailed information about age, sex, rearing history and previous experience with being proposer/receiver, meaning participation in Jensen et al.'s study (2007a) where chimpanzees played the ultimatum game with a different apparatus, quantities and without an analysis of multiple reciprocal trials..

Table 3.
Age, sex, rearing history and participation in Jensen et al (2007a) *of the sample.*

	Age	Sex	Rearing	Participation in Jensen et al (2007a)
Group A				
Lome	12	M	Mother	Yes
Lobo	9	M	Mother	No
Kara	8	F	Mother	No
Kofi	8	M	Mother	No
Group B				
Alex	13	M	Hand	No
Jahaga	21	F	Mother	No

Facilities

All the subjects were housed in large outdoor (Group A: 4.000m^2; Group B: 1.400m^2, see Figure 1) or indoor enclosures (Group A: 430m^2; Group B: 175m2, see Figure 2) depending on the weather conditions. Both enclosures contained a variety of vegetation (trees, shrubs, grass) as well as materials and structures for climbing, foraging and resting (see Figure 2). During the test we used two testing rooms, one for each social group. In each room the apes faced each other in different rooms (approximately

[29] As this was not the aim of our study and we only had two subjects, these results are not reported.

3x2x3m each, see Figures 3 and 4). There was a rectangular booth area (subject's side 0.8m x lateral 1m) in between both rooms where the experimenter placed the apparatus. During the game, subjects could only get access to the materials on the booth through a decision window allocated in their respective sides. They could not physically interact while playing, but they could do so between trials by moving to the crystal/mesh wall that separated both rooms (i.e. pass little objects from one room to another; touch their hands; climb upwards simultaneously).

Figure 1. Outdoor enclosure for chimpanzees (group A-right; group B-left).

Figure 2. Indoor enclosure for chimpanzees (group A).

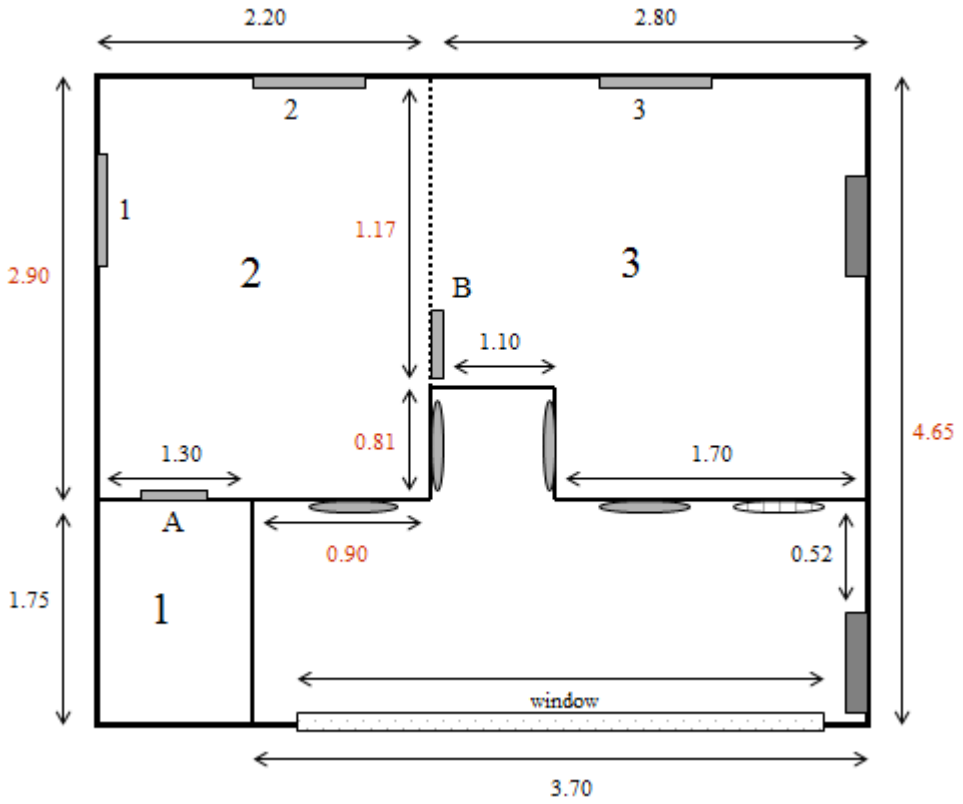


Figure 3. Experimental room for group A.

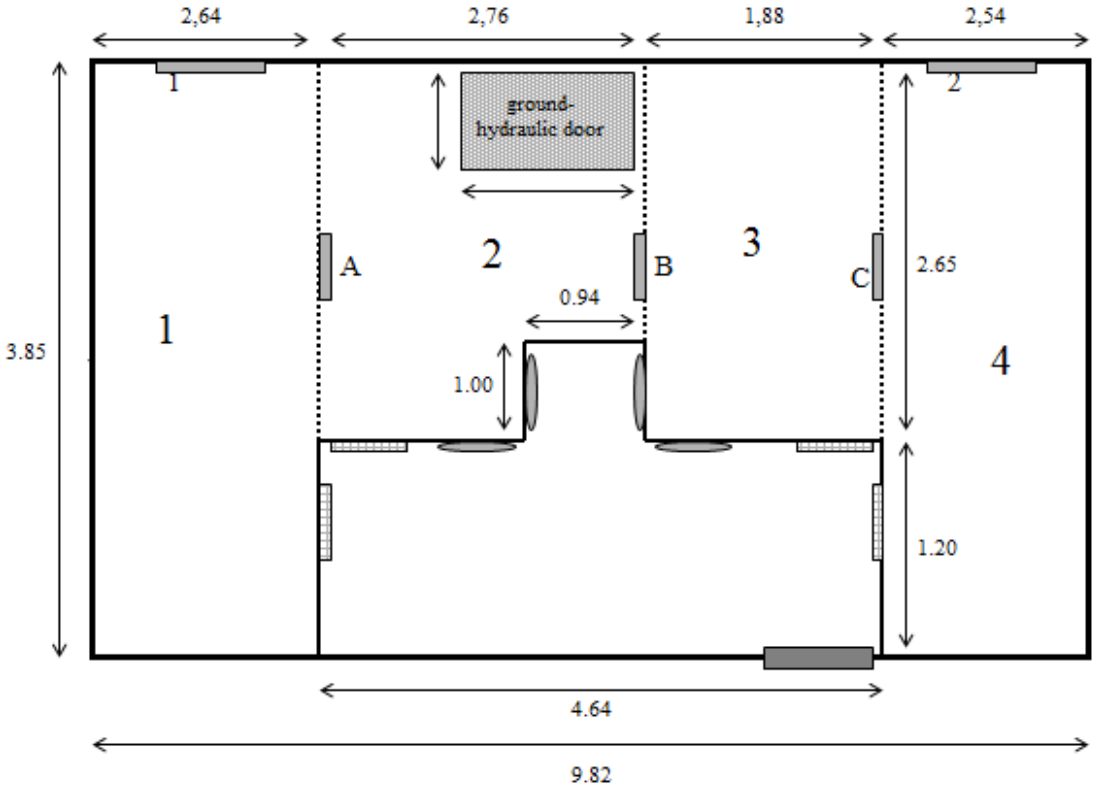


Figure 4. Experimental room for group B.

Study 1: materials

Apparatuses

We used two similar apparatuses adapted according to the game we played (DG or MUG). The apparatus for DG consisted of a hard PVC table (93 x 68 cm) with two parallel guide rails running from one subject's side to their partner's. Each guide rail had a pair of trays on it (20 x 13 cm each), on the top of which there was a dish (10 cm diameter) fixed with Velcro where the rewards were placed. Each pair of trays represented one different option. Every trial consisted of two options and the proposer could only choose one. To do so, the proposer had access to two 28 cm ropes so that by pulling from one of them the trays in that rail would move in opposite directions: the closest tray to his side, and the furthest tray to the partner's. To make their choice, the subjects introduced their fingers through one of the two holes (8.5 cm diameter) of the Plexiglas decision window (68 x 35cm) fixed to their respective sides to get access to the ropes. A removable transparent piece (50 x 13cm) blocked these holes before the trial started to allow them seeing the options before choosing (see Figure 5).

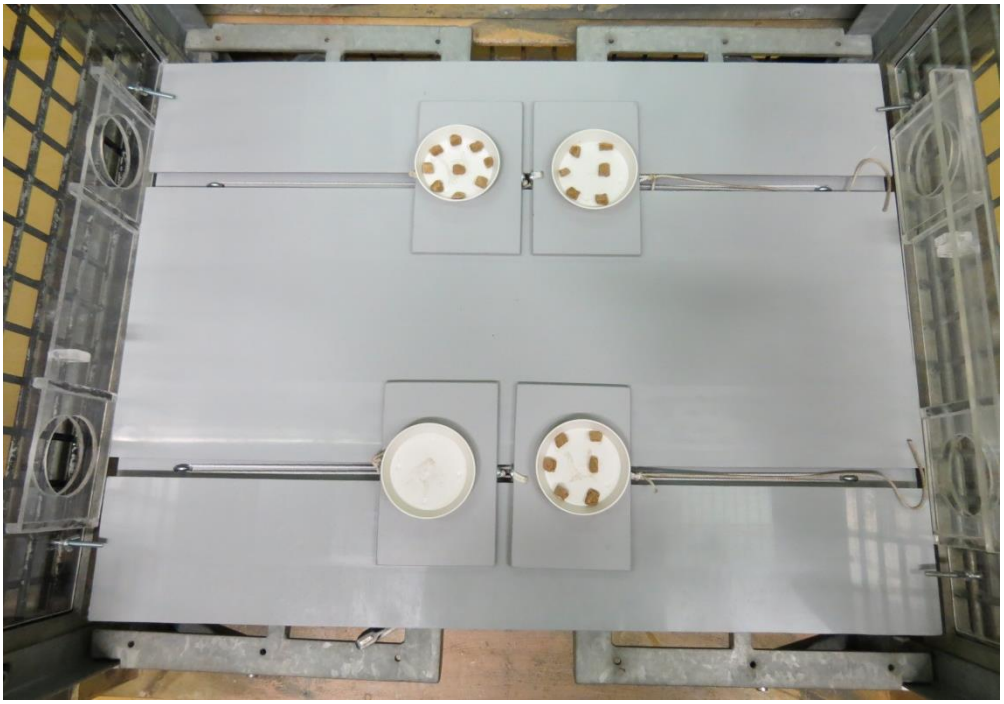


Figure 5. Top view of the DG apparatus. The PVC stands on two metal structures. In this photo, the proposer is on the right.

In MUG we used a similar apparatus to the previous one, consisting of a hard PVC table (90.5 x 67.5 cm), also with two parallel guide rails running from one subject's side to their partner's and each of them having a pair of trays on it (20 x 12.5 cm) with their respective dishes. The difference came from how the trays moved, as in this game the

respondent is active and therefore there are two choices (first proposer, then responder). By this token, the proposer had access to two 29 cm pieces of Velcro (shape "?" in the photo), so that by pulling from one of them the trays in that rail would move in opposite directions but in this case stopping at a certain distance from which the respondent had access to a handle (a looped Velcro fixed to the partner's tray) that he could pull to complete the shift of the trays and thus accept the offer. We used Velcro instead of ropes because the material allowed the handle to remain rigid after the trays had moved. We used the same transparent decision windows described above (see Figure 6).

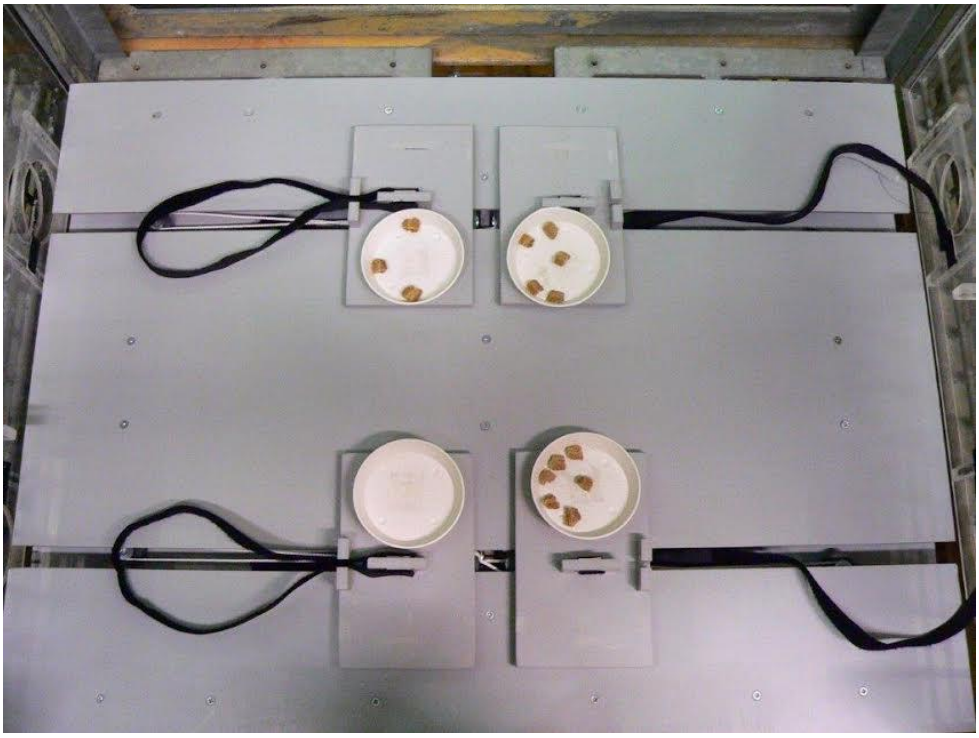


Figure 6. Top view of the MUG apparatus. The PVC stands on two metal structures. In this photo, the proposer is on the right.

In DG and MUG the experimenter used an occluder (a box 68 x 45 x 30cm with an upper hole, see DG and MUG Procedures, pages 91-92) to prevent subjects seeing the experimenter placing the rewards before the trial started.

Rewards

We used banana pellets or grapes, depending on the dyad's preference. The actual rewards were 1.5cm in diameter pieces of each (i.e. a third of a pellet; half of a grape). We took into account the maximum amount of food that apes can eat per day not to

interfere with their usual diet. The subjects showed interest for this food and found it appetizing.

Experimenter's material

The experimenter recorded the sessions with a digital camera on a tripod and coded live the data on a coding sheet. The apparatus rested still on two metal supports. The experimenter used a clamp to fasten the apparatus to those metal structures. The rewards were always inside a box, out of sight from the subjects.

Study 1: conditions

Conditions

The participants should be able to compare quantities accurately and prefer bigger quantities to play the MUG. In a pretest, chimpanzees showed no tendency to prefer 10 over to 8 pieces of food nor even 8 over 5 pieces of food. Therefore, we could not use with chimpanzees the same conditions we used with children in Study 2. Nevertheless, this limitation does not preclude further comparisons children-chimpanzees, because we changed the conditions in a way that IA (both disadvantageous and advantageous), frustration and cost to be generous were still present (see Figure 7).

Each trial consisted of two preassigned options of food. Following research on how great apes process quantities (Hanus & Call, 2007) we configured conditions to have higher differences and lower ratios between final outcomes to facilitate discrimination. As a result, we kept one default option (6/0) across four conditions controlling for cost to be generous (cost/no cost) and level of generosity[30] (low/high). The conditions were the following:

Table 4.
Experimental conditions that we used during the DG and MUG.

Name	Quantities	Cost for the proposer	Generosity to the responder
No Cost Small gift	6/0 and 6/3	No (6 vs. 6)	Small (0 vs. 3)
No Cost Large gift	6/0 and 6/9	No (6 vs. 6)	Large (0 vs. 9)
Cost Small gift	6/0 and 5/3	Yes (6 vs. 5)	Small (0 vs. 3)
Cost Large gift	6/0 and 5/9	Yes (6 vs. 5)	Large (0 vs. 9)

If we look at Table 4, in the No Cost conditions (first and second rows) the proposer could be generous with his partner at no cost (6 vs. 6 pieces of food), whereas in cost conditions (third and fourth rows) the proposer would incur a cost of one piece of food to do so (6 vs. 5 pieces of food). In the Small gift conditions (first and third row), the proposer could raise his partner's outcome up to 3 pieces of food, a quantity less than his own profit (i.e. 6 or 5 pieces of food). In the Large conditions (second and fourth

[30] We decided to shorten the level of generosity by using the word "gift".

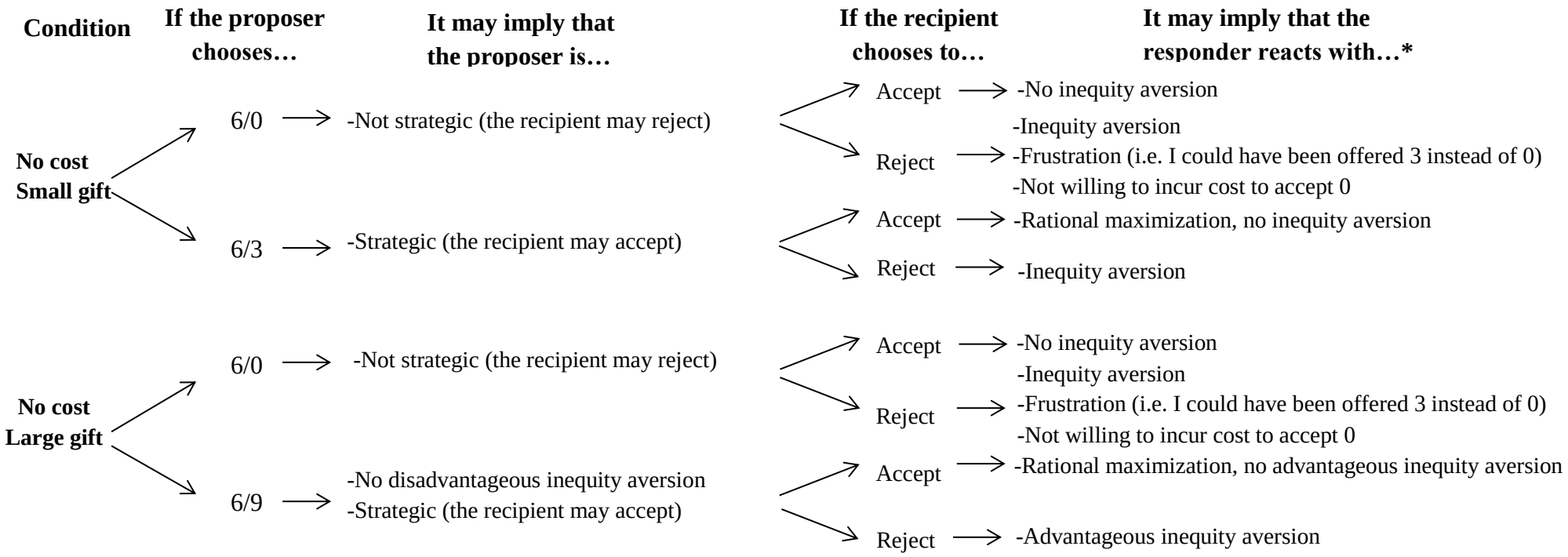
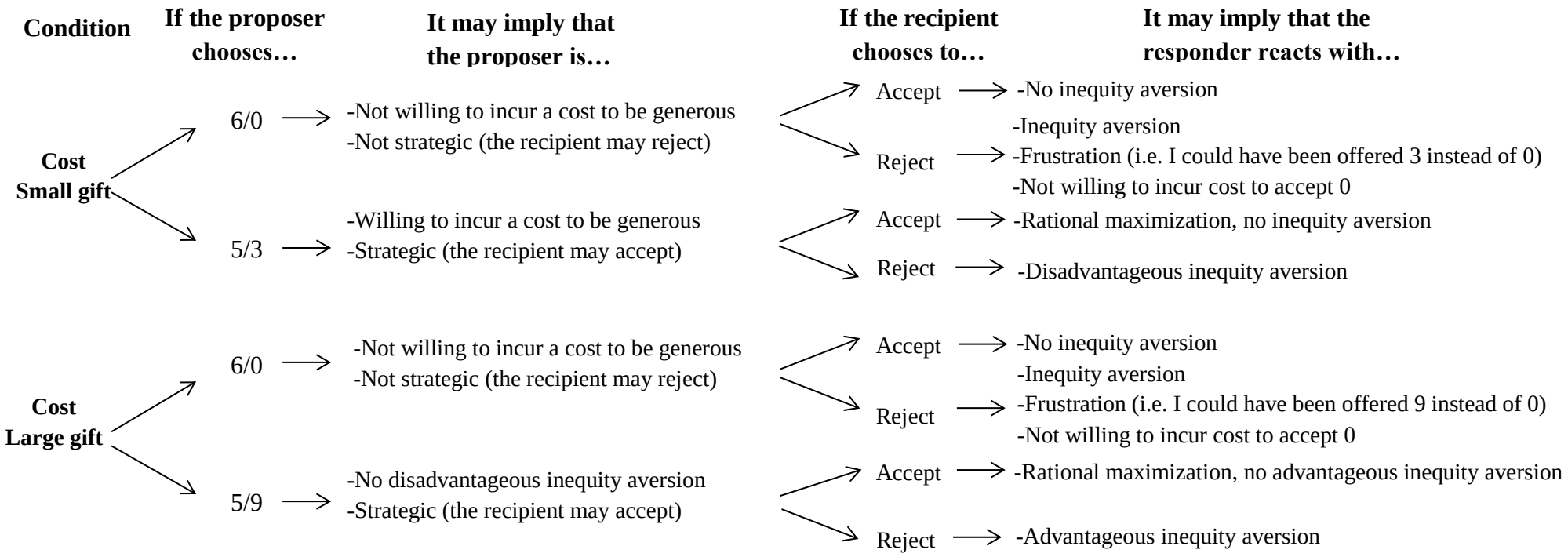


Figure 7. Complete outline of the different responses and interpretations in each condition of the MUG for both players. The rejections, if followed by a change of the proposer's behavior in successive trials, would work as functional punishment. We do not include this information in the last column because that may imply that we assign the responder an intention to promote cooperation and, as we have said earlier in the Introduction of this dissertation, intention is not always necessary for functional punishment and cooperation to arise. The scheme continues on next page.

* We have discarded "norms enforcing" since we are not sure that chimpanzees may have social norms in the first place.



(continuation from previous page)

Figure 7. Complete outline of the different responses and interpretations in each condition of the MUG for both players.

rows), the responder's outcome would surpass his/her own raising up to 9 pieces of food. Chimpanzee responders could experience disadvantageous IA in 6/0, 6/3 and 5/3 offers, advantageous IA in 6/9 and 5/9 offers and frustration in 6/0 offers. Chimpanzee proposers would show strategic behavior if they offered the non-zero option or incur a cost by choosing the 5/X options. In case this strategic behavior occurred during the development of the game, we should explore whether the responder's responses promoted it.

During the training (non-social trials, namely when subjects were tested individually), we used some of these conditions in order to meet the criteria we had to assure a full understanding of the conditions, but others varied (see Table 5). Generally, for subjects to understand the games, they had to show ability to (1) differentiate quantities at play and (2) make decisions according to how the food was distributed to maximize their outcome (i.e. understand that the quantities in the closest dishes will be for the proposer and the quantities in the furthest dishes will be for the partner in the opposite room). In order to achieve this, we played sessions with door open and sessions with door closed. Door open meant that the subject could move from one room (the proposer's side) to the other (the responder's side). Door closed meant that the subject could not move between rooms.

Table 5.
Conditions used during the training sessions by state of the door (open/closed).

Door Open	Door Closed
6/0 and **6/3**	6/0 and 6/3
0/0 and **0/3**	6/0 and 6/9
6/0 and **5/3**	**6/0** and 5/3
6/0 and **5/9**	**6/0** and 5/0

Note: the options in bold led the subjects to maximize their outcomes. There are no options in bold in the no cost conditions in door closed since the payoff was the same (6 vs. 6).

On the one hand, in Door open sessions the subjects could experience the consequences of their choices, namely what the partner would find during the tests. Given that they had access to the responder's room, we expected that the subjects maximized their outcome by choosing the options leading to more food in total. If we look at the first column in Table 5, this would imply choosing 6/3, 0/3, 5/3 and 5/9 respectively. Those conditions help subjects to learn to check the options before choosing and avoid natural impulses to pick always the highest quantity on their side (6/0) as well as to understand that the responder could obtain different payoffs depending on their choice from the

proposer's side. To make the need of paying attention to the responder's side more ostentatious we had a different condition from the conditions used during the tests (i.e. 0/0 and 0/3).

There was no learning effect (i.e. choosing X is always the maximizing option) because in Door closed sessions the maximizing options were different (i.e. in 6/0 and 5/3 condition choosing 5/3 is the most maximizing option when the door is open, but it is 6/0 when the door is closed). Also, Door closed sessions informed us about the subjects' discrimination of quantities (6 vs 5) and whether they considered the loss of 1 piece of food as costly and not desirable (third and fourth rows in Table 5). Finally, the condition 6/0 and 6/3 showed whether apes had a basal preference to pull from options with food on both sides, what we called the "just in case" (i.e. pulling from that option even after having understood that when the door is closed they cannot obtain food on the responder's side, just in case the experimenter, as the zookeepers in their daily routine, gives them something). Ideally, when the door is closed we should expect choices at chance in the no cost conditions and, if the subjects are prosocial, this tendency might change with the presence of a partner during tests (i.e. changing to a preference for 6/3).

In the post-trainings we used the same conditions as in the trainings (door open and door closed). However only two conditions from the training coincided with the conditions in the test (no cost and small gift: 6/0 and 6/3 and cost and small gift: 6/0 and 5/3). This is because we did not want to prolong the number of training trials, which could led to reduced motivation, and we preferred to lose some training-test comparison in some conditions (i.e. no cost and large gift: 6/0 and 6/9 and cost and large gift: 6/0 and 5/9) but gaining a solid confirmation of the subjects' understanding that otherwise would have been impossible (i.e. quantities' differentiation with 6/0 and 5/0; consequences of his choices with 0/0 and 0/3). Therefore, in the Results section, we will analyze the no cost-small gift and cost-small gift conditions in the training-postraining (only conditions common in door open/door close) and training-tests (only conditions in common) comparison and all the conditions in the DG/MUG comparisons.

Study 1: procedure

Procedure

The apes were not food deprived and they had water *ad libitum*. They participated voluntarily: they were called by their name and entered in the experimental room. If they did not participate within 10 minutes or if they showed signs of anxiety (i.e. there were screams outside due to a fight) the test was stopped and the subject was released. Two trained experimenters conducted the sessions (Nereida Bueno-Guerra and África de las Heras) during 2014-2016 (approx. 11 months with no more than 3 months without testing between testing periods). We followed an ABACA model (training (A), tests (B, C), post-training (A)). We controlled for game order (two dyads played DG/MUG; two dyads MUG/DG). The order of the sessions varied as it is shown in Table 6.

Table 6.
Order of sessions for ABACA model depending on which test the dyad played first.

DG first	UG first
Training – DG apparatus (8 sessions)	Training – DG apparatus (8 sessions)
DG test (8 sessions)	Training – UG apparatus (2 sessions)
Post-training – DG apparatus (8 sessions)	MUG test (8 sessions)
Training with MUG apparatus (2 sessions)	Post-training – DG apparatus (8 sessions)
MUG test (8 sessions)	DG test (8 sessions)
Post-training– DG apparatus (8 sessions)	Post-training – DG apparatus (8 sessions)

Training

During the training chimpanzees played individually to (1) differentiate quantities at play; (2) know they could only make one choice out of two options per trial; (3) understand how food was distributed (closest dishes for the proposer, furthest dishes for the partner in the opposite room) and (4) understand the role of their partner in every game (not able to respond in DG; able to pull in return in MUG). To do so, every subject played 8 sessions of 12 trials each: half sessions had the door in between rooms open and half closed. We counterbalanced the room where the actors played; the sides where the option of each condition was placed and the order of presentation of the conditions, for which we had eight different randomized models (each condition was played three times per session during no more than two consecutive trials). Each subject played the same randomized model across all the training and postrainings sessions. As I mentioned in the Conditions section, some conditions were different from those in the test in order to confirm their understanding of the task/quantities involved (see Table 4).

During the training we used the DG apparatus for 8 sessions (half door open + half door closed), however to understand the MUG procedure the subjects played with the MUG apparatus during 2 sessions right before starting the MUG test (see Table 6). In the MUG Training, the door was always open given the need to understand that accepting an offer entailed the partner pulling from the handle made of Velcro on his respective side. In the DG training we coded whether the subjects maximized their outcome whereas in the MUG training we also scored whether they pulled from the Velcro. The criteria to pass the training was to maximize 80% of the time in Door open and Door close (DG apparatus-sessions) and to choose the maximizing options plus pull from the Velcro 10 out of 12 trials in two consecutive sessions (MUG apparatus-sessions). All subjects passed without the need for extra sessions.

All the training sessions had the same procedure: the experimenter informed the keeper in which room the subject should be placed (right or left, counterbalanced); baited the dishes with food according to the randomized model for that particular subject while being out of sight thanks to the use of an occluder; placed the ropes/Velcro strips next to each hole of the decision window; removed the occluder; let the chimpanzee 10 seconds to see the options before taking the decision window out and moved the rope/Velcro from the other option away immediately after the chimpanzee had chosen one (to make the subjects understand that they could only choose one out of the two options). Once the subject had moved the dishes, the experimenter helped him/her to retrieve the food from the dish. If there was no food in a chosen dish, the experimenter showed the empty dish to the subject and placed it again on the platform. If the door was closed, the experimenter only fed the subject with the food on his/her side and did not touch the food on the responder's side. In the MUG sessions, if the subject did not pull from the rope within 15 seconds, the experimenter moved the dishes to the original position, unreachable, and started preparing the next trial. After every trial, the experimenter put the trays again in the initial position and the procedure started over.

Test

Each dyad played MUG and DG. Two dyads played MUG/DG (Kara+Kofi and Alex+Jahaga) and two dyads played DG/MUG (Lome+Lobo and Lobo+Kofi).

Each test consisted of 8 sessions with 6 reciprocal trials each (12 trials in total per session). In reciprocal trials the subjects played the same condition twice but

exchanging their roles, thereby each subject played the same condition first as proposer and then as responder or the other way around. The procedure was very similar to the training, except that after every trial the experimenter changed the position of the ropes/Velcros so that from proposer's position passed to responder's positions and viceversa. We counterbalanced in which room the subject was placed (right/left); which player started first; the presentation of the conditions (each condition was played no more than twice per session) and the sides where each option of a given condition was placed (right/left).

Every trial started with the experimenter placing the food out of sight from the participants (with the help of an occluder). After that, she opened the decision window of the responder, the occluder and when the ape was facing the apparatus, she also opened the decision window of the proposer. The ape chose one option and 15 seconds were counted for the respondent to pull (MUG) or both received the food through the mesh (DG and MUG if accepted). In case there was a rejection (15s without respondent pulling), the experimenter started to prepare the next trial. Thus, the intervention of the experimenter was minimal: she stayed back during the game and only helped chimpanzees getting access to the food through the mesh after the choice was made. We can be sure that the chimpanzees understood that they were playing with their partner instead of the experimenter because they used communicative with the conspecific and did not direct any behavior towards the experimenter. See next page for the initial configuration of the apparatuses (DG, Figure 8 and MUG, Figure 9); page 91 for the detailed DG procedure and page 92 for the detailed MUG procedure.

Postraining

After each test we ran the post-training (i.e. we ran 2 per subject). They played again the same eight sessions of twelve trials each (half door open, half closed). Thus we could explore whether their performance remained stable during non-social conditions and whether it was different from the performance in social conditions (tests). The procedure was the same as in the training.

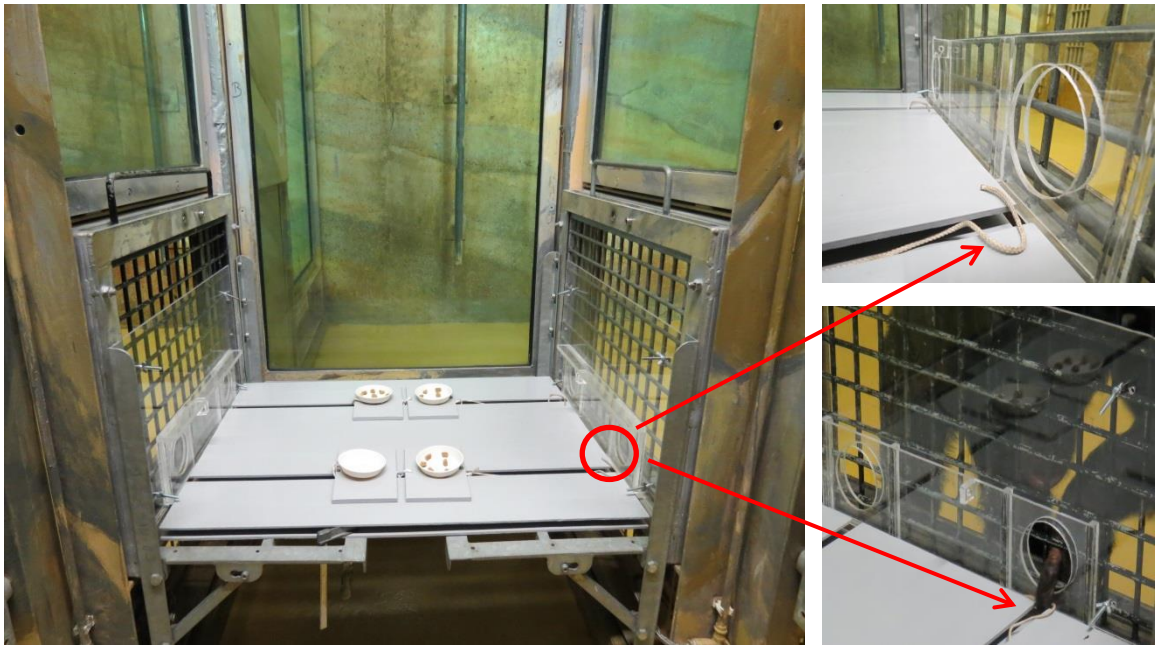


Figure 8. The DG apparatus as it was disposed during the experiment. In this photo, the proposer is on the right. The option displayed is 6/0 (foreground) and 6/9 (background). There are two ropes, one attached to each option, for the proposer to make an offer. The end of the ropes is blended next to the proposer's window (as it is displayed in the maximized photo in the margin) to allow the subject gathering the chosen option with his/her finger. When the proposer pulls from one rope the dishes of that option move in opposite directions along the trails, which are the dark horizontal lines on the platform. See procedure on page 91.

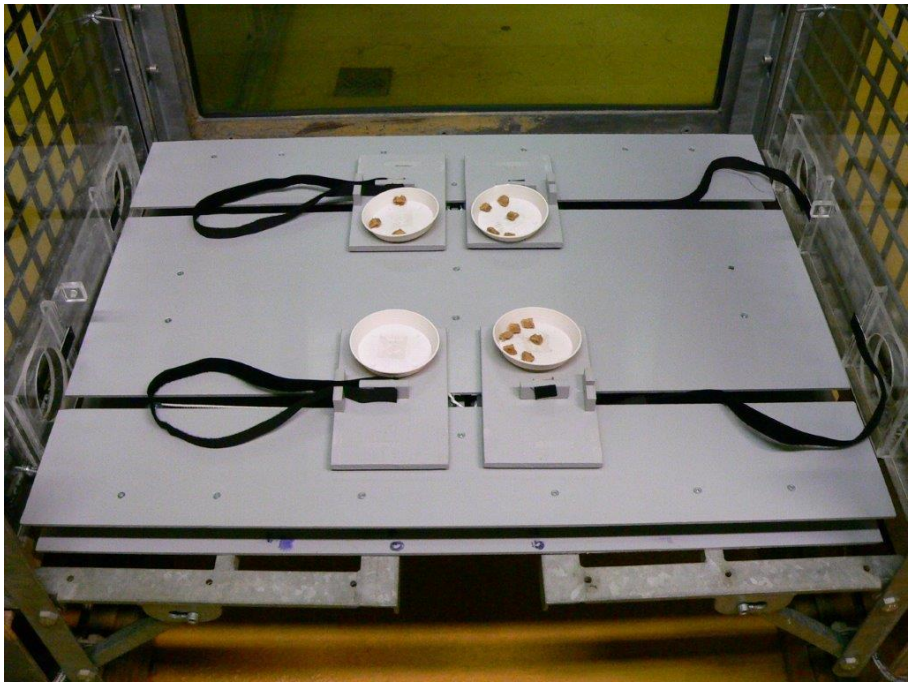
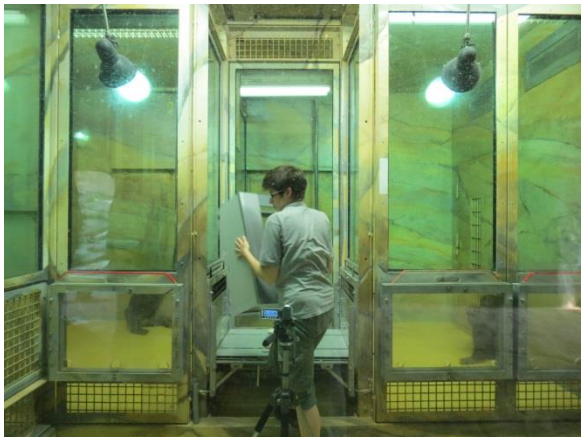
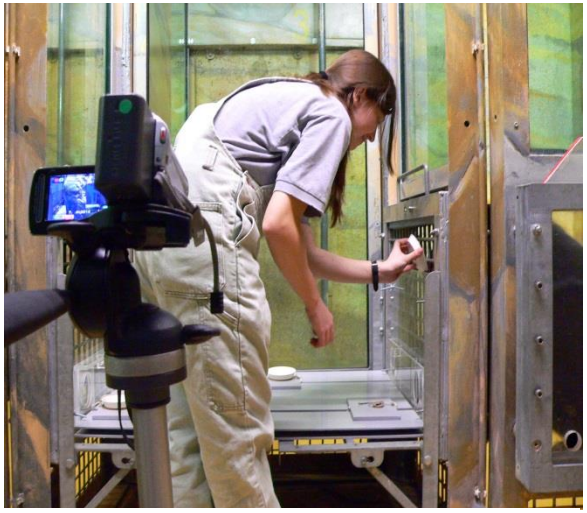


Figure 9. The MUG apparatus as it was disposed during the experiment. In this photo, the proposer is on the right (Velcros in "?" shape) and the responder is on the left (Velcros like "handles"). The option displayed is 6/0 (foreground) and 5/3 (background). When the proposer makes an offer, the dishes in that rail stop halfway so that the handles became accessible to the responder. See procedure on page 92.

Detailed procedure for the DG:

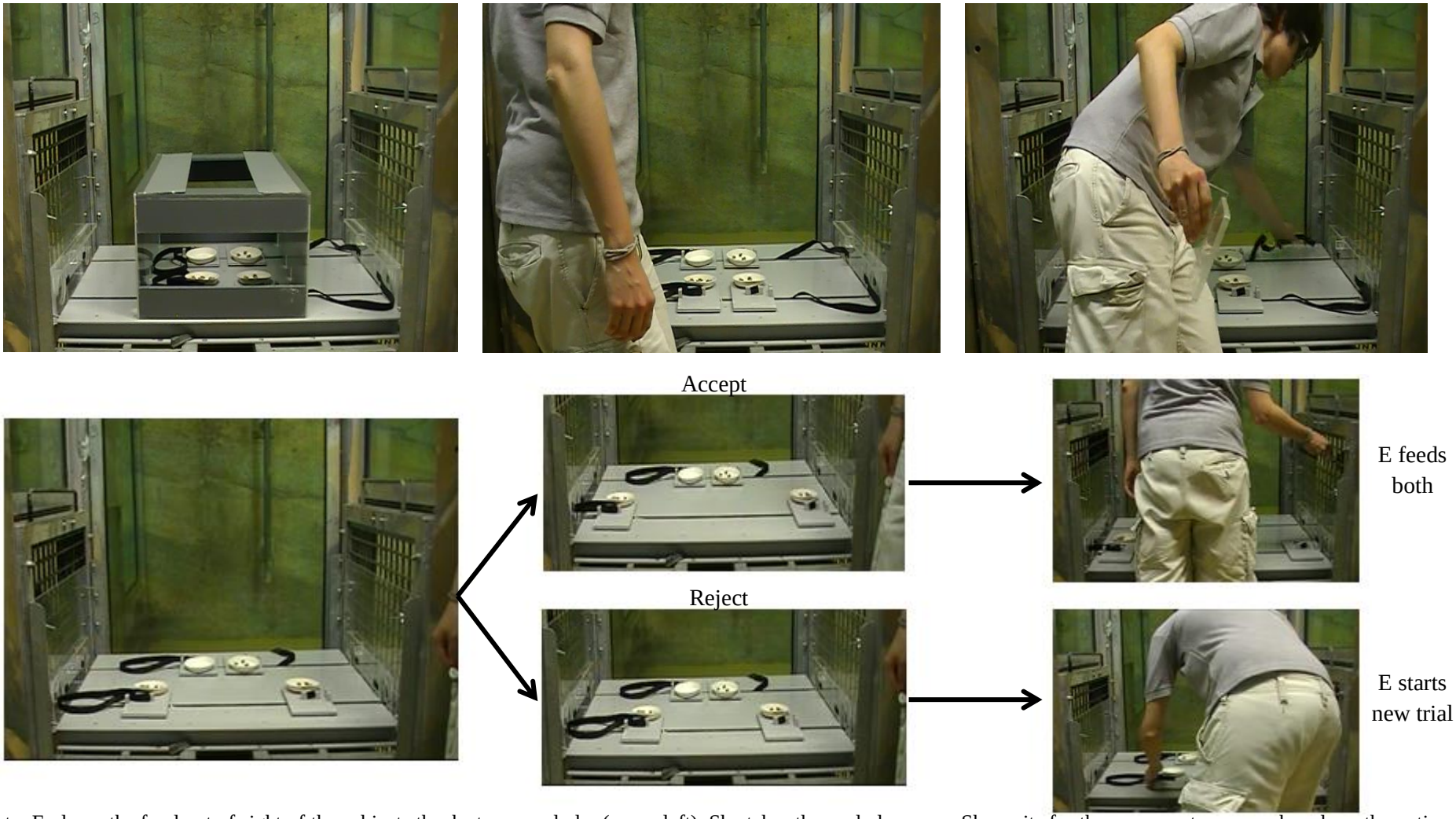







Note: E places the food out of sight of the subjects thanks to an occluder (upper left). She takes the occluder away (upper middle). She waits for the proposer to approach and see the options (upper right). The proposer pulls from one option. E has only taken the rope from the alternative option away and waits. (bottom left). The trays move forward (bottom middle). E helps subjects feeding from their respective dished (bottom right).

Detailed procedure for the MUG:



Note: E places the food out of sight of the subjects thanks to an occluder (upper left). She takes the occluder away. She waits for the proposer to approach and see the options (upper middle). She takes the decision window away. The proposer makes his choice pulling from one Velcro. E takes the alternative Velcro away and waits (upper right). The proposer has pulled from one option. A handle is available for the recipient during 15 seconds (bottom left). The partner can accept the offer by pulling from that handle, completing the movement of the trials (bottom-middle-up) or reject by not pulling (bottom-middle-down). In case the partner accepted, E helps both participants feeding (bottom-right-up). In case he rejected, E initiates a new trial (bottom-right-down).

Analysis[31]

We used a Generalized Linear Mixed Model (GLMM; Baayen, 2008) with binomial error structure and logit link function to analyze subjects' choices. When subjects delivered food to the opposite side (prosocial) we scored 1, when they chose the option that delivered food only to their side (selfish) we scored 0. We divided this section into: training and post-training comparison (non-social); tests comparison; tests separately and non-social and tests comparison. The latter analysis includes: differences by type of game, reciprocity, rejections and communicative gestures.

To keep type I error rate at the nominal level of 5%, we included all possible random slopes components (game, cost, and gift, and all 2-way interactions within subject; and for the tests also within partner and dyad) but not the correlation parameters between random intercepts and random slopes terms (Barr, Levy, Scheepers, & Tily, 2013; Schielzeth & Forstmeier, 2009). In the model comparing the non-social condition to the tests, we included subject and dyad as random effects (but not partner, given that there was no partner in the non-social control). We entered virtual dyads in the non-social condition to be able to include the random slope component of cost within dyad.

As an overall test of the effect of the predictor variables we compared the full model with a null model lacking the fixed effects and interactions but comprising the same random effects structure as the full model (Forstmeier & Schielzeth, 2011) using a likelihood ratio test (Dobson & Barnett, 2008). P values for the individual effects were based on likelihood ratio tests comparing the full with respective reduced models ((Barr et al., 2013); R function drop1). The model was fitted in R (R Core Team, 2016) using the function lmer of the R package lme4 (Bates, Mächler, Bolker, & Walker, 2015).

We assessed model stability by comparing the estimates derived by a model based on all data with those obtained from models with the levels of the random effects excluded one at a time. The models were stable with regard to the effects of game, gift, and cost. Over dispersion was no issue (dispersion parameters: training and post-training model: 1.00; tests: 0.98; non-social vs. tests: 1.00).

[31] I very much appreciate the help of PhD Christoph Völter with the analysis.

Study 1:
results

Non-social (training and post-training)

We first ran a GLMM with the nonsocial data to examine whether apes understood the set-up, namely if they were sensitive to the relevant pieces of information such as the state of the door, the amount of food on their side (labelled as "cost") and the food on the opposite side (labelled as "gift"). In the first model, we just analyzed the conditions that were common in both open and close door trainings (i.e. 6/0 vs. 5/3 and 6/0 vs. 6/3) and we included cost (yes/no), the state of the door (open /closed), session, trial number, training phase (training, post-training 1, post-training 2), the interaction between cost and door state as well as subject ID as random effect. We also examined the effect of the amount of food at the opposite side (i.e. 3 and 9, gift) on apes' performance.

Overall, the GLMM with trial number, session, training phase, door state, cost on the subject's side, size of the reward at the opposing side (gift), and the 3-way interaction between door state, cost, and gift was significant compared to the null model (likelihood ratio test: χ^2=88.81, df=10, p<0.001; see Figure 10 and Table 7). More specifically, we found a significant 3-way interaction between door, cost, and gift (χ^2=3.86, df=1, p=0.049). We did not find a significant effect of trial number (χ^2=1.28, df=1, p=0.258), session (χ^2=0.004, df=1, p=0.950), or training phase (χ^2=0.05, df=1, p=0.831) on performance.

Figure 10 shows the percentages of choices. From left to right, in cost conditions when the door was closed the maximizing choice was choosing 6/0 instead of the alternative option (5/0 or 5/3), since choosing the alternative implied losing 1 piece of food (6 vs. 5). Indeed, chimpanzees chose 6/0 around 70%-80% of the time. Importantly, even though they saw that the option 5/3 had food on the other side (i.e. 3) they consistently preferred 6/0 instead. If we continue with the Figure 10 we can see that in cost conditions when the door was open the maximizing choices were the alternative options (6 vs. 8 (5+3) vs. 14 (5+9)). Indeed, chimpanzees chose the alternative options around 60%-85% of the time. In no cost conditions when the door was closed there should be no preference for the default or the alternative options, since both of them provided 6 pieces of food. Chimpanzees did not show chance level but a tendency to prefer the alternative options: 6/3 (around 60%) and 6/9 (around 75%). Finally, in no cost conditions when the door was open the maximizing choices were the alternative options, and chimpanzees clearly chose them: 6/3 (around 79%) and 0/3 (around 97%)

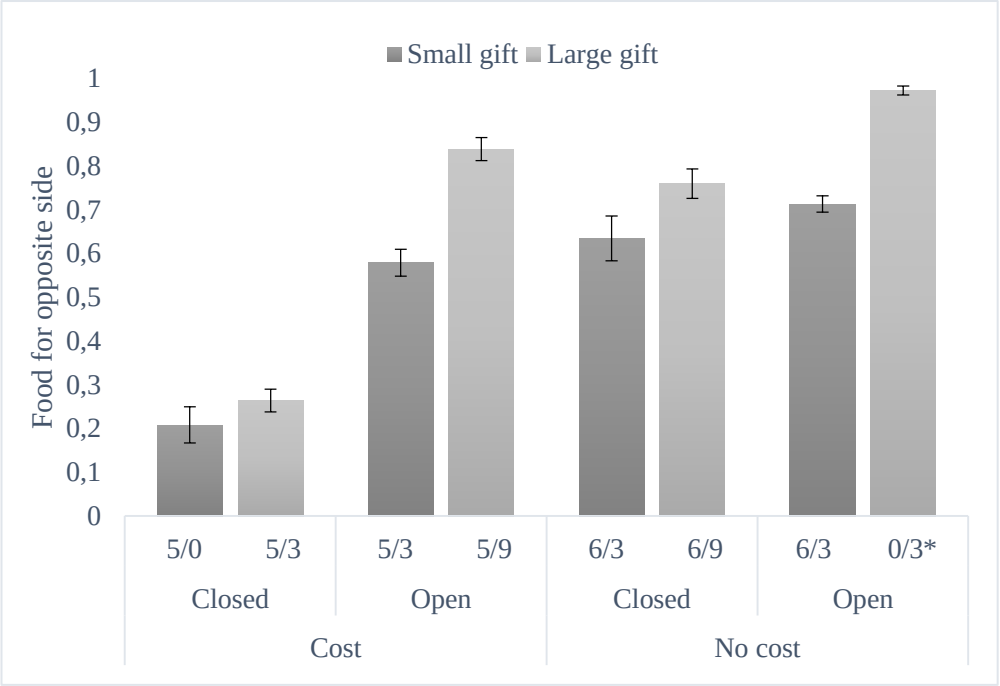


Figure 10. Training performance (mean ± SE) as a function of condition and door state. The alternative option was 6/0 in all conditions except for the 0/3* condition in which the alternative was 0/0.

Table 7.
Output of GLMM of the no-partner training.

	Estimate	SE	z	P
(Intercept)	-1.017	0.155	-6.578	<0.001
Door	2.671	0.241	11.092	<0.001
Cost	2.182	0.239	9.147	<0.001
Gift	-0.384	0.244	-1.575	0.115
Session	-0.004	0.058	-0.062	0.950
Trial	-0.069	0.059	-1.178	0.239
Training phase	0.012	0.058	0.214	0.831
Door:Cost	-0.260	0.505	-0.515	0.607
Door:Gift	-0.933	0.327	-2.854	0.004
Cost:Gift	-0.240	0.315	-0.760	0.447
Door:Cost:Gift	-1.117	0.589	-1.897	0.058

Post-hoc test revealed a significant interaction between door state and the size of the gift irrespective of whether there was a cost at the subjects' side (χ^2=8.50, df=1, p=0.004) or not (χ^2=20.81, df=1, p<0.001). Concretely, subjects' preference for the large gift at the opposite side was more pronounced when the door was open than when it was closed (i.e. subjects preferred to take the alternative with 9 pieces of food on the other side when the door was open and they could get it than when the door was closed and they

could not obtain that amount). Subjects significantly preferred the large gift over a smaller one when the door was open irrespective of whether there was a cost for the subject (χ^2=17.41, df=1, p<0.001) or not (χ^2=26.64, df=1, p<0.001) (i.e. when the door was open and therefore the alternative options (5/9, 5/3, 6/3) were the maximizing choices, subjects had a preference for 5/9, namely 14 pieces of food, rather than 5/3 and 6/3, namely 8 and 9 pieces of food, respectively). Finally, when the door was closed, subjects significantly preferred the larger gift to a smaller one when there was no cost (χ^2=7.06, df=1, p=0.008) (i.e. they preferred 6/9 and 6/3 over 6/0). This shows a baseline tendency to choose "prosocial" options when there is no cost for them. However, this did not happen when there was a cost at the subject's side (χ^2=1.96, df=1, p=0.161) (i.e. they preferred 6/0 over 5/0 and 5/3). Therefore, chimpanzees care very much about cost even though the difference was only 1 piece of food.

Taken together these results established that subjects paid attention to the quantities on their side (i.e. cost: yes/no) and on their partner's side (i.e. gift: small/large) as well as understood the effect of pulling each of the ropes on each side in the apparatus since they adapted their choices to the door state to maximize their outcome. Moreover, they also had certain tendency for prosocial options in no cost conditions. The latter should be faced to social conditions (tests) to see whether the presence of a partner has an influence per se or eventually has it after rejections (cooperation through functional punishment).

Non-social and tests (MUG and DG)

To examine whether the presence of a partner affected subjects' responses, we compared the non-social conditions with the social games. We analyzed the conditions that were common in MUG, DG and nonsocial door closed (i.e. 6/0 vs. 5/3 and 6/0 vs. 6/3). We ran another model in which we included game (MUG, DG, door-closed non-social), cost for the subject (yes, no), and the 2-way interaction between these factors as well as subject, partner, and dyad as random effect. Since our analyses had shown that subjects' responses in all non-social trials remained unchanged throughout testing, we pooled these data to create a more solid baseline and simplify analyses and called them "control".

Figure 11 presents the percentage of trials in which chimpanzees selected the option that delivered food to their partner's side as a function of game (control, DG, MUG,) and

cost at the subject's side. A GLMM with game and cost was significant compared to the null model (χ^2= 25.66, df=3, p<0.001). Proposers were more willing to deliver food to the other side when there was no cost for them (-1.40 ± 0.15, χ^2= 19.67, df=1, p<0.001). Moreover, we found a significant effect of game (χ^2= 6.20, df=2, p=0.045). More specifically, subjects delivered more food to the other side in DG compared to the control (-0.43 ± 0.19, χ^2= 4.93, df=1, p=0.026). In contrast, we did not find that difference between the MUG and the control (-0.06 ± 0.19, χ^2= 0.11, df=1, p=0.736). Consequently, the chimpanzees chose significantly more prosocial options at no cost in the DG compared to when they played alone.

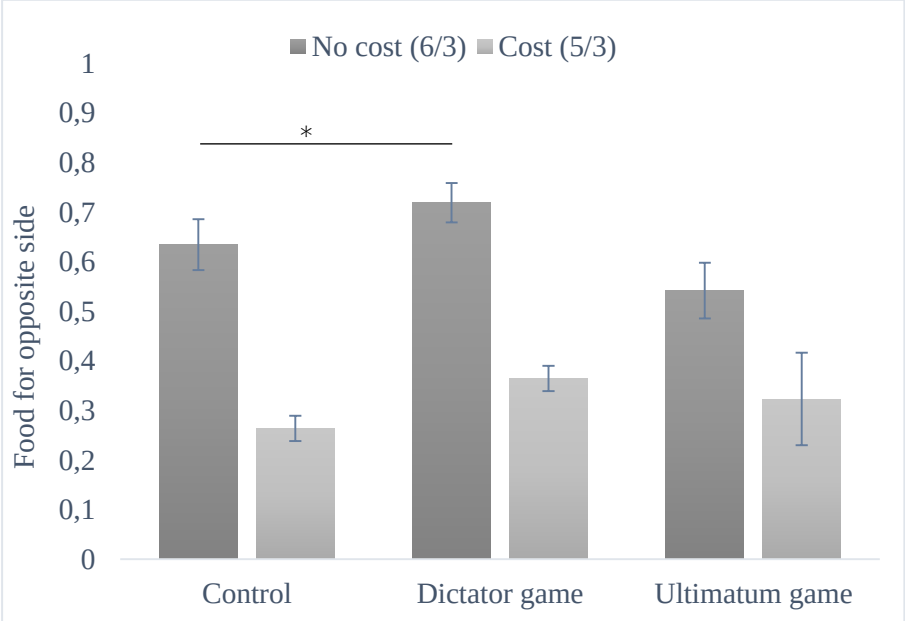


Figure 11. Percentage of trials (mean ± SE) in which proposers selected the option that delivered food to their partner's side as a function of game (control, DG, MUG) and cost for the proposer.

Tests separately (MUG and DG)

We also analyzed the two games separately by including cost (yes, no), gift (small, large), session, and trial number (but no interactions) in the models. In the DG, proposers were significantly more willing to deliver food to the partner when there was no cost for them compared to when there was a cost (-1.35 ± 0.27, χ^2=9.35, df=1, p=0.002). In contrast, there was no significant effect of the size of the gift given to the partner (0.40 ± 0.22, χ^2=2.85, df=1, p=0.092), session (-0.11 ± 0.11, χ^2=0.95, df=1, p=0.331), or trial number (0.03 ± 0.11, χ^2=0.06, df=1, p=0.803) on performance. Different from the DG, in the MUG proposers were significantly more willing to deliver

food when the gift for the partner was large than small (0.93 ± 0.24, χ^2=7.52, df=1, p=0.006). Moreover, proposers got less inclined to deliver food over the course of sessions (-0.28 ± 0.11, χ^2=4.96, df=1, p=0.026). Also, proposers chose similarly both costly and no costly options (-0.79 ± 0.46, χ^2=2.47, df=1, p=0.116) and there was no significant effect of trial number (-0.06 ± 0.11, χ^2=0.28, df=1, p=0.598). Moreover, the responders accepted all offers above zero whereas zero offers were accepted in 58.3 ± 7.1% of trials. We examined factors that may have affected rejection rates. We included the zero-offer trials in a GLMM with the factors session, cost, gift, and the interaction between cost and gift was not significant compared to the null-model (χ^2=2.26, df=4, p=0.687), neither was a reduced without the interaction (χ^2=1.76, df=3, p=0.624).

Hence, the chimpanzee proposers were paying special attention to the cost they incurred in the DG and to the amount of the food delivered for the partner in the MUG. In the MUG responders accepted zero-offers a little bit more than half of the time, coinciding with a tendency in the proposers to decrease generosity over sessions.

Test comparison (MUG vs. DG)

For the analysis of the social games, we included trial number, session, game (MUG, DG), cost for the subject (yes, no), and the amount of food for the partner (3, 9), all interactions between game, cost, and gift as well as subject, partner, and dyad as random effects. We considered all conditions. This model was significant compared to the null model (χ^2=27.50, df=8, p<0.001). We found no significant interactions between game and cost (0.57 ± 0.52, χ^2=1.11, df=1, p=0.291), game and gift (0.58 ± 0.32, χ^2= 3.27, df=1, p=0.073), or cost and gift (0.29 ± 0.32, χ^2=0.81, df=1, p=0.367). A reduced model without the 2-way interactions was significant compared to the null model (χ^2=24.46, df=5, p<0.001).

Subjects were more willing to deliver food to the partner when there was no cost (-1.13 ± 0.16, χ^2=13.28, df=1, p<0.001) and when the gift for the partner was large (0.65 ± 0.20, χ^2=5.32, df=1, p=0.021). Moreover the likelihood to deliver food to the partner decreased over sessions (-0.19 ± 0.08, χ^2=3.91, df=1, p=0.048). There was no significant difference between the games (-0.19 ± .020, χ^2=0.83, df=1, p=0.361). Neither was the main effect of trial number (-0.003 ± 0.08, χ^2=0.002, df=1, p=0.967).

Figure 12 presents the percentage of trials in which chimpanzees selected the option that delivered food to their partner's side as a function of game (DG, MUG), gift for the partner (3 vs. 9) and cost at the subject's side (5 vs. 6). Taken together these results show that chimpanzees played DG and MUG similarly, namely they preferred not to incur costs to be prosocial regardless of the active/passive role of their partner but they were at least generous with the amount of the food they could deliver.

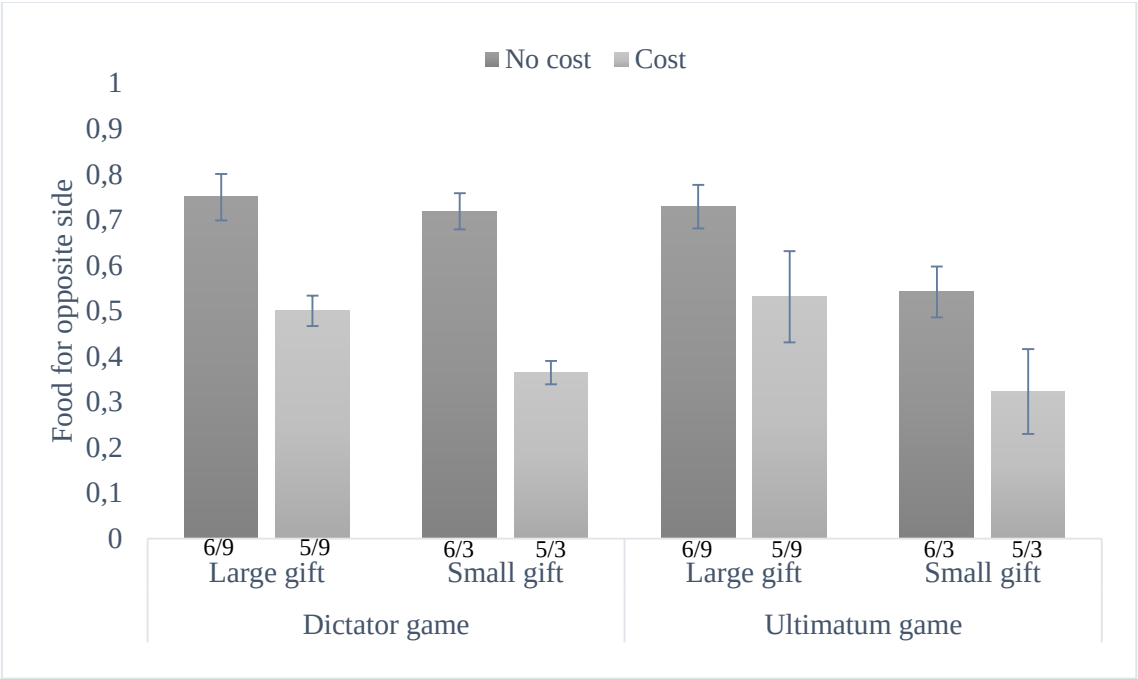


Figure 12. Percentage of trials (mean ± SE) in which chimpanzees preferred the option that delivered food to their partner's side over the default 6/0 option as a function of game (DG, MUG), size of the gift for the partner (large, small), and cost at the subject's side (cost, no cost).

Reciprocity[33]

We analyzed whether receiving a reward as responder in the previous trial affected the same subject when switching to proposer in the following trial by increasing his willingness to provide the partner with food. For the reciprocity analysis, we included game (MUG, DG), cost (yes, no), gift (small, large), sessions, whether subjects received food in the previous trial (yes, no), the 2-way interactions between these factors and game, gift, and cost in the current trial, and as random effects subject, partner, and dyad. Overall, this GLMM was significant compared to the null model (χ^2=24.37, df=8,

[33] We only report positive reciprocity (cooperation after cooperation) because there were no rejections of options different from zero. Nevertheless we explored potential negative reciprocity (defeating after defeating) and it was absent. Besides, rejections did not make proposers' offer more generously. At the contrary, the tendency to offer generously decreased over sessions.

p=0.002). To evaluate potential main effects we removed the non-significant interactions from the model. The reduced model was significant compared to the null model (χ^2=25.35, df=5, p<0.001; see Table 8). We found a marginally significant effect of session (χ^2=3.81, df=1, p=0.051) with decreasing food offers for the partner over the course of the experiment. There were no significant interactions between food received in previous trial and game (χ^2=0.22, df=1, p=0.636), cost (χ^2=0.34, df=1, p=0.560), or gift (χ^2=1.34, df=1, p=0.247).

Table 8.
Output of the reduced GLMM (once excluding non-significant interactions) analyzing the effect of the previous trial on subjects' willingness to provide the partner with food in the current trial.

	Estimate	SE	z	p
(Intercept)	0.556	0.207	2.688	0.007
Previous choice	0.181	0.184	0.982	0.326
Game	-0.166	0.200	-0.828	0.407
Gift	0.773	0.237	3.261	0.001
Cost	-1.177	0.175	-6.719	<0.001
Session	-0.180	0.082	-2.193	0.028

At the individual level, only one chimpanzee was significantly more likely to offer food to his partner after having received food in the previous trial (Fisher's exact test: p<0.05, see Figure 13).

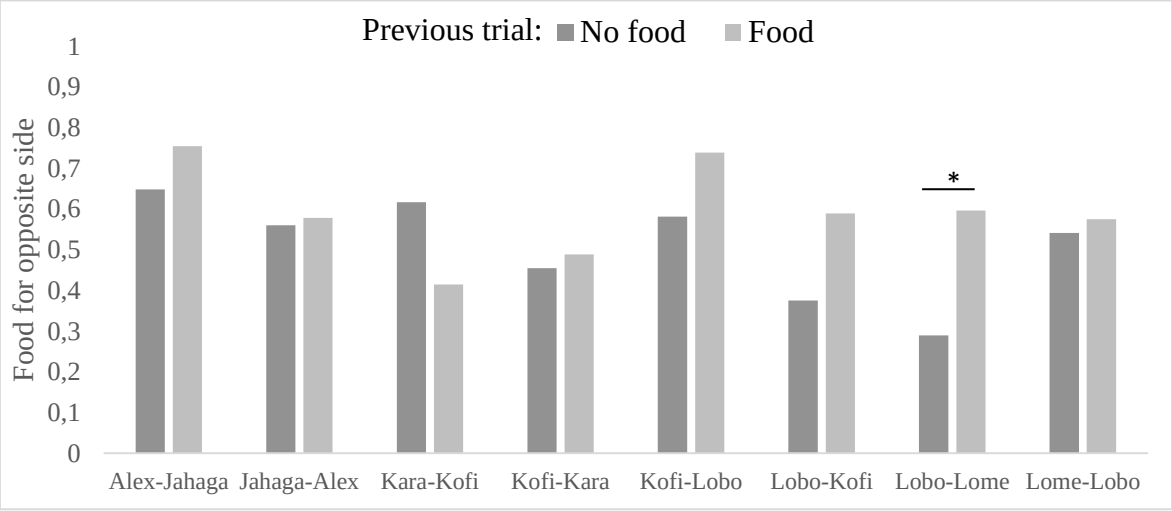


Figure 13. Individual performance: the effect of receiving food in the previous trial on subjects' willingness to provide food for the partner in the current trial.

Communicative attempts

Finally, we also examined the effect of communicative attempts between proposer and responder. We coded "interaction", when the subjects touched or passed objects to each other through the mesh away from the apparatus, and "pointing", if the subjects placed their index finger or their hand through the decision window for more than 3s addressed to the maximizing distribution for them. We analyzed the two responses separately.

To examine whether the chimpanzees perform any of these responses at different rates in each social game we used the frequency of these responses as dependent variable and included game (MUG, DG) as factor in a GLMM and subject, partner, and dyad as random effects. Moreover, as pointing and interaction could enhance the probability of the proposer to deliver more food (i.e. choosing 6/3 instead of 6/0) or to incur a cost (i.e. choosing 5/3 instead of 6/0), we analyzed whether these communication attempts were related to the proposer's choice. Therefore we included the factors communicative response (pointing, interaction), game (MUG, DG), and cost (yes, no) and the 2-way interactions between the communicative response and game and cost, respectively, together with the random effects subject, partner, and dyad in a GLMM.

Interaction between proposer and recipient. We analyzed only the data of dyads in which direct interactions between proposer and recipient occurred (6/8 dyads). Overall, interactions between participants occurred only 34 times (5.9% of all trials). We first analyzed whether the frequency of interactions was affected by the game. Therefore, we used interaction frequency as dependent variable and included game as factor in a GLMM. There was a significant effect of game (estimate ± SE: -2.85 ± 1.63, χ^2=4.41, df=1, p=0.036) indicating that the frequency of interactions was higher in MUG than DG.

Table 9.
Output of GLMM analyzing the effect of game on the frequency of the interactions between proposer and recipient.

	Estimate	SE	z	p
(Intercept)	-4.884	1.196	-4.083	<0.001
Game	2.602	1.324	1.964	0.050

We observed interactions between proposer and responder in the ultimatum game 29 times, 17 times initiated by the proposer and 12 times by the recipient. We analyzed the

effect of interaction between proposer and responder on the proposer's choices in the MUG. We included cost, social interaction, and the interaction between these two factors in the model. The model was not significant when compared to the null model (χ^2=5.58, df=3, p=0.134). Therefore, we removed the interaction to evaluate potential main effects. The model approached the significance level when compared to the null model (χ^2=5.67, df=2, p=0.059). There was a significant effect of cost (χ^2=5.15, df=1, p=0.023) but not of the social interaction between participants (χ^2=0.54, df=1, p=0.464) on the proposer's performance.

Table 10.
Output of GLMM analyzing the effect of game, cost, and interaction on the proposer's choice.

	Estimate	SE	z	p
(Intercept)	0.609	0.182	3.350	0.001
Interaction	-0.495	0.667	-0.742	0.458
Cost	-1.155	0.384	-3.008	0.003

Pointing. We analyzed only the data of dyads in which pointing occurred (7/8 dyads). Pointing gestures were always produced by the recipient except for a single instance (99.0%). We first analyzed whether the frequency of pointing was affected by the game. Therefore, we used pointing as dependent variable and included game as factor (and subject, partner, and dyad as random effects) in a GLMM. Subjects tended to produce more pointing gestures in MUG (mean ± SE: 22.0 ± 7.8%) than in DG (8.6 ± 4.4%; estimate ± SE:-1.98±0.92, χ^2=3.40, df=1, p=0.065).

We then analyzed the effect of pointing by the recipient on the proposer's performance. We included game, cost, pointing, and the interaction between pointing and game and cost, respectively, in the model. The model was significant compared to the null model (χ^2=16.08, df=5, p=0.007). However, the interactions between pointing and game (estimate ± SE: -0.72 ± 0.72, χ^2=0.95, df=5, p=0.329) or between pointing and cost (-0.37 ± 0.75, χ^2=0.27, df=1, p=0.600) were not significant. Therefore, we removed the interaction to evaluate potential main effects. The reduced model was significant compared to the null model (χ^2=14.91, df=3, p=0.002; see Table 11). Aside from an effect of cost (χ^2=11.48, df=1, p<0.001), surprisingly, we found that when the recipients pointed the proposer tended to deliver food less often (χ^2=3.19, df=1, p=0.074).

Table 11.
Output of GLMM analyzing the effect of game, cost, and pointing on the proposer's choice.

	Estimate	SE	z	p
(Intercept)	1.054	0.161	6.528	<0.001
Pointing	-1.074	0.518	-2.074	0.038
Game	0.050	0.268	0.186	0.853
Cost	-1.138	0.196	-5.817	<0.001

In the next page there is a summary of the results obtained in Study 1 (Table 12).

Table 12.
Summary of the results in iterated DG/MUG experiment with chimpanzees.

Concept	Key points
Training	Chimpanzees maximized according to the state of the door (open/close). They preferred 6 over 5 pieces of food. They foresaw the maximizing option before moving to the opposite room.
Training, Post-training (non-social)	Chimpanzees maximized consistently every time they played alone. We pooled the data.
Non-social, DG, MUG	The presence of a conspecific was only significant when playing the DG (proposers were more generous).
DG	Proposers took care of cost. They significantly chose more prosocial options in no cost conditions. Proposers chose indistinctly small and large gifts.
MUG	Proposers took care of gift. They chose significantly more large gifts than small gifts. Proposers chose indistinctly cost and no cost options. The tendency to be generous decreased across sessions. Responders rejected approximately 58% of zero offers. They never rejected non-zero offers.
DG, MUG	Proposers played DG and MUG similarly. Proposers preferred not to incur costs to be prosocial. Proposers preferred large gifts over small gifts. The tendency to be generous decreased across sessions
Reciprocity	There were no signs of positive or negative reciprocity except for Lobo. When he played with Lome, he responded accordingly to what Lome had offered in the previous trial. However, Lome did not reciprocate Lobo.
Interactions	There were more interactions between proposer and responder in MUG. Interactions do not influence the proposer's next offer.
Pointing	There were more pointings in MUG. Pointing does not influence the partner's action (no better offers/no more acceptances)

Study 1:

discussion

Discussion

This is the first time that an iterated MUG/DG protocol with an ABACA design in chimpanzees is conducted. Our results show that chimpanzees kept consistent preferences when playing alone and that they understood the apparatus and the procedure of each game, namely they paid attention to the quantities on their side, the opposite side and to the consequences of choosing between the two options available. A direct comparison between the offers made in the MUG and the DG revealed no significant differences. This in principle means that any rejection by the responders in the MUG did not affect the proposer's behavior. Interestingly, the proposers seemed sensitive to different factors in each game. In the MUG proposers tended to choose the option with more food for the partner (i.e. x/9 over x/3), whereas in the DG proposers tended to choose the option with less cost for them regardless of the amount they were giving to the partner (i.e. 6/x over 5/x). This could be interpreted as if chimpanzees were playing with certain strategy in the MUG. However, only in the DG the proposers' offers were higher than those observed when they played alone, thus compatible with the existence of intrinsic (although non costly) prosociality. By contrast, in the UG chimpanzee proposers did not make different choices than when they played alone. Indeed, they significantly decreased their donations over time. We did not find evidence of IA since responders never rejected offers bigger than zero, but only approximately half of 10/0 options. The decrement of donations might be explained by an accumulated frustration after 10/0 rejections, which might have made proposers more selfish, unwilling to reward the partner after zero-outcome trials. This is revealing, because, contrary to what humans usually do, the partners' punishment (i.e. rejection) led chimpanzee proposers to offer less generously rather than change their strategy and promote cooperation. This is compatible with chimpanzees lacking of strategic behavior (and therefore, with no clear evidence of second-order IA). Moreover, it seems that responders are rational maximizers and their punishment in 0-options does not work as a functional punishment because it is not perceived as a "warning for cooperation" in the proposers. Finally, we found no evidence of positive reciprocation.

Unlike humans, chimpanzees behaved as rational maximizers, invariably accepting offers bigger than zero, what takes advantageous or disadvantageous IA out of the non-human primates' picture, at least while bargaining. This is a very robust finding that has now been replicated in four different studies (this one plus Jensen et al., 2007a; Kaiser

et al., 2012; Proctor et al., 2013a). On the other hand, similarly to Jensen et al.'s study (2007a) more than a half of the zero offers were accepted, which might be hard to reconcile with rational maximization. It has been argued that such high acceptance rate might imply lack of understanding of the task (Brosnan, 2013) however, we have shown that all subjects in our study understood the contingencies of the game and their responses remained stable across the non-social conditions. Another explanation could be found in an excessive time delay for rejections that could have fostered acceptances to make a new trial start (Brosnan, 2013; Smith & Silberberg, 2010), but this is unlikely since we reduced the waiting up to 15 seconds (lower than 60s in (Jensen et al., 2007a) and 30s in (Proctor et al., 2013)). Thus, as Henrich and Silk claimed (2013), it might be that in a game where both accepting and rejecting a zero option entailed unavoidably zero outcome as well, rejections occurred at chance, thus in accordance with our results and (Jensen et al., 2007a), and still compatible with rational maximization.

With regard to proposers, it seems that there are signs of a tendency to provide food to conspecifics in game contexts since the chimpanzees proposers chose prosocial options more in DG than when they played alone. In fact, such prosocial offers resemble those made by humans in the same game when the recipient has no power, and are also in line with the change of preferences to offer more in social rather than non-social conditions of Proctor et al.'s study (2013a). However, from a classical economics point of view, Proctor et al.'s result (2013a) was puzzling because proposers offered more than expected in an MUG even though recipients never rejected any offers. One explanation was prosociality or anticipatory avoidance of conflict (Brosnan & de Waal, 2014) but some methodological concerns made these explanations contentious. Briefly stated, the absence of 0-option made apes' understanding of rejection difficult to sustain; the exchange of tokens through a begging human could force the recipients to consistently accept all offers and the lack of a non-social condition made impossible to know whether the proposers' responses were influenced by their partner's presence. In the current study we do not replicate that result but find that chimps behave prosocially when the role of their partner is passive, supported by the contrast between DG and the non-social condition. This is very important, because it implies that chimpanzees do not rely on functional punishment when playing iterated MUG. Proposers facing a rejection of a selfish option (and therefore a resulting zero outcome) would not be willing to reward the partner with food in a future trial and persist in offering less and less food.

This would be similar to keep on punishing the recipient for rejecting instead of changing to reward the recipient to make him more willing to accept. If that was the case, there would be no signs of second-order inequity aversion in chimpanzee proposers after the recipients' refusals, contrary to previous interpretations (Proctor et al., 2013), but a lack of strategic behavior. Indeed, the fact that chimpanzees were more prosocial in DG rather than in UG is likely due to the effect of the half 0-rejections on the proposers. We believe that in case those rejections had not existed, the prosocial tendency would have been similar between both games. Therefore, contrary to expected, rejections in chimpanzees provoke frustration and selfishness, not cooperation.

Moreover, we provide evidence of a calculated prosociality by the contrast between cost and no cost conditions showing that chimpanzees did not choose prosocial options indiscriminately but preferentially when there was no cost involved. However, the same proposers did not offer more or at least similarly in the MUG than the DG, as it could have been expected to avoid rejections. This is potentially due to the recipients' behavior. While human proposers face high risk of rejection, chimpanzee proposers do not. Chimpanzee responders had a high acceptance rate (they never reject non-zero and only approximately half of zero options), which turned any selfish offer likely to be accepted at least half of the time. Together with a frustration effect, this low rejection rate would justify the significant decrement of prosocial choices in MUG.

Despite their large methodological differences, the four studies conducted so far (Jensen et al., 2007; Kaiser et al., 2012; Proctor et al., 2013 and the present one) have consistently shown that human and non-human primates seem to differ when they play MUG and DG, suggesting a divergent evolutionary pathway in the consideration of IA and functional punishment.

Study 2: MUG with children

State of the art[34]

When sharing resources, a proposer can opt for a purely selfish perspective (the largest share for me). However, when reaching an agreement between a proposer and a recipient is necessary, the proposers should include their partners' outcome preferences and the recipients should consider outcomes with the interplay of the donor's intentions (Blake, McAuliffe, & Warneken, 2014; Falk & Fischbacher, 2006). Recent years have seen great interest in tracing the development of fairness in children particularly in relation to strategic behavior (proposers) and functional spite (recipients) (eg. Blake, Piovesan, Montinari, Warneken, & Gino, 2015; Camerer, 2003; Sally & Hill, 2006). The MUG is an experimental methodology that can help to test both and trace how they develop with age. We used the classical version (Falk et al., 2003): children faced a default option (8/2) pitted against an alternative option: Fair condition (8/2 and 5/5); Hyperfair condition (8/2 and 2/8); No Choice condition (8/2 and 8/2) and Hyperunfair condition (8/2 and 10/0). For our purposes, on the one hand, we will analyze strategic behavior in the proposers. This will inform us whether, in absence of prior rejections, children proposers are able to offer the option that is likely to be accepted (i.e. they offer 5/5in Fair and 8/2 in Hyperunfair). In case they were able to consistently do that, it would mean that they are "ready" to stablish cooperation from the very beginning. On other hand, we will analyze functional spite (i.e. rejections) in responders. The existence of different conditions will inform us about whether children apply functional spite specially when the proposer's intention was selfish (for example, if responders reject 8/2 in Fair but they do not in Hyperunfair means that functional spite relies on the proposer's intention). In this sense, the condition No Choice is very informative, because if children reject there, this will mean that functional spite is not mainly based on the proposer's intention but in any other motive to reject (i.e. IA, frustration, norms enforced). Importantly, the relevance of conducting one-shot anonymous MUG is that rejections, unlike an iterated game, will not provide any future outcomes for the responder, as the functional punishment would, and therefore it is an example of revenge.

Compared to the large body of evidence available in adults (see Güth & Kocher, 2014, for an updated review), there are still few studies that have investigated in detail

[34] This section is adapted from the Introduction of the study that we already published (Bueno-Guerra et al., 2016)

children's responses in the UG and MUG and how they change with age (eg. Fehr, Bernhardt, & Rockenbach, 2008; Kail & Cavanaugh, 2004). Moreover, many of these studies have found contradictory results. For example, Harbaugh and colleagues (2003) found that proposers in the UG offered better outcomes as they grew older whereas Murnighan and Saxon (1998) found the opposite. Similar mixed results on age and sex have been found in recent contributions in adults (eg. Eckel & Grossman, 2008; García-Gallego, Georgantzís, & Jaramillo-Gutiérrez, 2012; Saad & Gill, 2001; Solnick, 2001). In children, sex differences have not been explored (eg. Güroğlu et al., 2009; Sutter, 2007; Wittig et al., 2013). However, it is conceivable that such differences may exist given that boys and girls differ in their attitudes toward game play (Vogelsang, Jensen, Kirschner, Tennie, & Tomasello, 2014). Boys are more task-oriented (prioritize rewards) and competitive than girls who seem to focus more on interpersonal proximity (prioritize cooperation) (Maccoby, 2002), which could reflect sex differences in how individuals progress through the establishment of cooperation between partners. Jaffee and Hyde (2000) concluded that sex differences may be found in some age ranges only, specifically during adolescence, with girls showing slightly higher levels of prosocial reasoning than boys (Beutel & Johnson, 2004). Consequently, our understanding of whether children are able to offer "cooperatively" a priori, without the need of a previous rejection, with special attention to differences between boys and girls, remains unclear.

To our knowledge, only four MUG studies have been conducted with children and adolescents (Gummerum & Chu, 2014; Güroğlu et al., 2009; Sutter, 2007; Wittig et al., 2013). These studies do not offer a clear picture of the development of strategic behavior or functional spite. For example, children of similar ages in two studies (Güroğlu et al., 2009; Sutter, 2007) made vastly different offers (three times lower). It is conceivable that differences in the age distribution in the respective samples may have been responsible for this outcome (Güroğlu et al., 2009 only included 9-year-olds whereas Sutter considered children from 7 to 10 years of age). Moreover, the earliest record of disadvantageous IA is placed at 5 years of age (Wittig et al., 2013). Blake et al. (2015) placed it at 4 years of age but their study was outcome-based, without the interplay of intentions. However, the rejections of the default option never happened in the Hyperunfair condition, different from the next reported age in MUG, 9 years old, where rejections of the default option happened in all conditions. Therefore, it remains

unclear how proposers adapt their offers to the pattern of rejections as well as how functional spite varies across development.

The use of different methodologies may have substantially contributed to the mixed results available (see House, Henrich, Brosnan, & Silk, 2012 for a comprehensive review). Generally, two procedures have been proposed for playing MUG in the lab: the strategy-method and the direct-response-method. In the strategy-method, participants answer questions about their hypothetical offers and their thresholds of acceptance. The proposer is asked the maximum amount he would be willing to offer to a partner in case he had certain amount of resources, and written down. The responder is asked the minimum he will be willing to accept as an offer, and written down. Then the responses of both players are crossed to look for general tendencies within a population (i.e. What do proposers generally offer? What do responders generally accept?). As they do not interact with a real partner, this method is also called "cold play". Data are collected separately for proposers and responders and are used to find general tendencies within a population. Sometimes, the proposers' choices are recycled (for example, if they only composed 10% of the sample, as in Gummerum & Chu, 2014; or if there were no shows, as in Sutter, 2007). This is the most popular method across research (using computers: Gummerum & Chu, 2014; Güroğlu, van den Bos, & Crone, 2009; Sally & Hill, 2006; using paper and pencil: Leman, Keller, Takezawa, & Gummerum, 2009; Sutter, 2007), maybe because of its simplicity, low cost and sample size requirements (one subject can "play" -answer questions- both roles, thus doubling the number of participants). However, answering hypothetical questions about offers, acceptances and rejections may not activate the same cognitive and emotional processes as playing with another individual (Güth & Kocher, 2014; Sally & Hill, 2006). In contrast, "hot play" or the direct-response method places two subjects one in front of another (anonymity can be controlled) and allows them to play by taking their respective turns. This method is more similar to real life scenarios in which we directly interact with other people, but it also adds several personal-dependent factors that can interfere with the decisions if the situation is not anonymous or several methodological and pragmatic difficulties if the situation is anonymous (such as how to avoid visual clues or how to control children not to look through the occlude, see Procedure below). Data are collected on each dyad which allows a more refined and direct analysis of the individuals' choices. All MUG studies with children to date have used the strategy method except one that used the

direct-method (Wittig et al., 2013) using an apparatus with two shelves displaying the different amounts of stickers and pulling/pushing alternatives to propose and respond, respectively. The interesting fact is that the strategy and the direct-response methods have generated different data with regard to the responders. Concretely, there is a higher rejection rate of uneven offers under the direct-response method (for a comparison between MUG studies conducted with children see Supplementary Materials in Bueno-Guerra et al., 2016). Consequently, the presence of a social partner matters. However, as most MUG studies have used the strategy-method, it might happen that the majority of the conclusions regarding bargaining in children are biased or differ from real markets (Sally & Hill, 2006). That is why we opted for the direct-method both for humans and chimpanzees[35].

The difference between rewards played in the game and the rewards earned is another aspect that may affect children's responses, especially those of younger participants. Usually, people interact with other people about resources under clear and common terms. The most prominent example of this is price markets: there is fully accessible information about a product so that clients know that some particular object costs a particular amount of money. Negotiations are usually undertaken about some distinguishable resources and it is the conditions or the distributions which can vary. Indeed, if the information about the resources is not clear-cut, it is conceivable that uncertainty or too much cognitive load in trying to "do the maths" ended up interfering in the social interaction. Despite this argument might seem common sense, all available MUG/UG studies except for Wittig and colleagues' experiment (2013) have used a conversion procedure to translate the number of rewards obtained in the game into actual "take-home" rewards (e.g., if the participant earns 10 points during the game, he/she will take 2 euros home (Sutter, 2007)). To complicate comparisons across studies even more, each study has used a different conversion rule. What might have precluded past research to use a "what you see is what you get" procedure? Probably lack of funding to pay for all the rewards that a sufficient sample size might entail (eg. if all 150 participants played only one MUG condition and all the responders accepted, the

[35] The strategy-method as it has been explained does not seem easy to apply with chimpanzees (how can we ask them for desideratums such as how much would you be willing to offer/accept? The use of tokens in Proctor and colleagues' study (2013) explained in the Study 1's state of the art section, is a direct-response method because despite chimpanzees did not directly manipulate the food, they chose between two objects that led to two quantities, which is the same but just adding an intermediate step (the use of objects representing amounts).

available funding will necessarily be 1.500$. Moreover, if all the subjects passed through all the conditions this quantity would need to be much bigger, which is almost unaffordable for any research group for just one experiment). In our study we decided to ese the procedure to our young participants and follow the rule "what you see is what you get". We paid for the rewards thanks to the help of the Department of Psychology and Clinical Psychobiology.

Another aspect that affects responses, at least in adults, is whether games are played as 'one-shot' or as repeated interactions (Cooper & Dutcher, 2011; Lin & Sunder, 2001). Repeating the ultimatum game led adult participants to learn norms, deduce thresholds, and decline low offers, which does not happen when the game is played in a "one-shot" modality. Most studies with children have used repetitions of the game which, just like with adults, may have affected children's responses. Besides, for our purposes one-shot methodology was crucial to test revenge and therefore to prevent expectations for changes in the proposer's behavior. Precisely because of that our subjects played only one condition (eg. if we would have let them to play more, they could have harbored the idea that by rejecting once they could "teach" the partner what they expected as fair).

Anonymity was another important consideration. Personal features such as the name, social distance or the beauty of participants influence behavior in bargaining games (Charness & Gneezy, 2008; Marchetti, Castelli, Harlé, & Sanfey, 2011). Wittig and colleagues (2013) had used a non-anonymous test to enhance the truthfulness of the scenario. Although the participants belonged to different classes of the same kindergarten, this could still have affected their performance since children care about their reputation in front of people they know and share more resources (Engelmann, Over, Herrmann, & Tomasello, 2013). Therefore, we do not know whether children responses were affected by the visual access to their partner and we decided to achieve absolute anonymous interactions in our study. The study of what anonymity exactly is turns into a very ample subject. (eg. Having no visual clues? Absence of identity? Knowing the sex but not anything more from the partner?). The definition of anonymity and how anonymous environments affect our behavior is very well described in a purposefully anonymous research publication (Anonymous, 1998). In experimental research the impact that physical traits have on behavior has been called the "Proteous Effect", which alludes to the mythological deity Proteus, who could turn its physical appearance into anything he wanted. These effects can be related to sociocultural rules

or to more unspecific causes. As evidence of sociocultural rules, in Western adults there is a significant difference in the rate of rejections when bargaining depending on the sex of the partner, so that men tend to make higher offers for women and always accept the offers coming from them ("the chivalry effect") and women tend to reach an agreement between them ("the solidarity effect") whereas men do not ("male competition effect") (Eckel & Grossman, 2001; Saad & Gill, 2001). As evidence of changes triggered by unspecific causes, personal features such as the name, social distance or beauty of participants have also shown to produce significant differences in bargaining games (Charness & Gneezy, 2008; Marchetti et al., 2011). The influence of physical features is as striking as the fact that differences in height can provoke differences in bargaining (eg. "children who were one standard deviation taller than the mean proposed about half a token less than children who were one standard deviation shorter" ((Harbaugh et al., 2003, p. 17). With this overwhelming data on how non-anonymous interactions have an influence over bargaining behavior, we decided to prevent subjects from seeing and listening to each other. They did not know the sex or the name of their partner. The unique information available they had was that the partner was of their same age.

Table 13.
Summary of the most frequent shortcomings in UG procedures and how we tried to solve them in the present study.

	Implementation
Factors	
Age	Early and late childhood
Sex	Girls and boys
Methods	
Procedure	Direct method, real scenario
Payoff	What you see is what you get
Anonymity	Opaque apparatus
One-shot	One interaction per dyad

Our study was aimed at resolving the existing discrepancies in the literature by targeting critical ages (6 and 10 years of age) and addressing the various methodological concerns that we referred to. We implemented the MUG using a direct-method one-shot direct-reward delivery that guaranteed anonymity (see Table 13) with 6- and 10-year-old children as a way to provide data about the development of strategic behavior and functional spite. Wittig and colleagues' study (2013) is not vulnerable to most of those shortcomings that we mentioned earlier but it is limited to 5 years of age. We studied 6-year-old children because this is the age when they begin formal schooling and are

confronted with resource sharing on a daily basis –an activity that might increase the likelihood of showing certain strategic behavior and functional spite. Furthermore, we studied 10-year-olds because this is the age when children consistently punish transgressors based on others' intentions (Helwig, Zelazo, & Wilson, 2001; Zelazo, Helwig, & Lau, 1996). Note that previous studies have not directly compared these two ages that mark the beginning and end of primary education, respectively.

Our first hypothesis tested whether strategic behavior was present at 6 or 10 years of age: a preference for the option different from 8/2 when it benefited their partner (5/5 in Fair; 2/8 in Hyperfair) and the avoidance when it did not (10/0 in Hyperunfair). To that end we analyzed how proposers of the two age groups distributed their offers across conditions. We also examined whether the offers matched the pattern of rejections, namely if what was offered was usually accepted by responders, since this would mean that the proposers could anticipate the responders' expectations (i.e. they are ready to cooperate). We predicted that 6-year-olds would not show a homogeneous pattern of offers whereas this will be likely to exist already at 10 years of age. Our second hypothesis tested sex differences in strategic behavior. It is interesting to know whether a particular sex is "readier" than the other to engage in cooperation without the previous need of rejections. Here we compared the offers of boys and girls. We predicted that girls would show more strategic behavior than boys because girls develop moral reasoning earlier than boys. Such a difference should be particularly marked in our older participants. Our third hypothesis tested the presence of functional spite. For this analysis, we focused on the responder's rejections. In order to provide information about the existence of IA, we considered that the 8/2 rejections in Hyperunfair and No Choice conditions inform us about disadvantageous IA because there was no better option available. The rejections of 2/8 in Hyperfair informed us about the presence of advantageous IA. Finally, we explored whether functional spite was based on the proposer's intentions (comparison between rejections of both options in Hyperunfair and between 8/2 option in Fair and No Choice). We expected to find advantageous IA in both ages, since at 5 years was already present somehow (Wittig et al, 2013) whereas we expected disadvantageous IA to increase with age due to the increase of social comparison. Finally, we expected both age groups to show functional spite, based on both outcomes and intentions.

Study 2: sample

Children tested

I tested 338 first-grade children (mean age = 6.56 years, SD = 0.36), paired in 169 dyads (175 girls: 88 proposers and 87 responders; 163 boys: 81 proposers and 82 responders) and 140 fifth-grade children (mean age = 10.49 years, SD = 0.32), paired in 70 dyads (77 girls: 39 proposers and 38 responders; 63 boys: 31 proposers and 32 responders). We randomized the sample so that there were a similar number of boys and girls playing each role and the same number of dyads playing each condition (a full description of the distribution can be seen in Table 15). Ninety-four percent of the participants were Spanish; the rest were of several other nationalities (Russian, Moroccan, Pakistani, Chinese, and Latin American). The experimenter spoke either Spanish or Catalan during the experiment, depending on the child's preference.

Participating schools

In Table 14 we show the participating schools from which we recruited our sample. There were four public and two private elementary schools:

Table 14.
School, type (public/private), city and age distribution of the sample.

School name	Type	City	N	
			6-y-olds	10-y-olds
25 Setembre	Public	Rubí	108	40
Torre de la Llebre	Public	Rubí	36	20
Institut Lloret de Mar	Public	Lloret de Mar	28	16
Ramón Llull	Public	Barcelona	30	-
Arrels Esperança	Private	Badalona	98	48
Arrels Blanquerna	Private	Badalona	38	16
		Total	338	140

As it can be seen in Table 16, the sociodemographic origins of our sample was homogeneous and representative of the Spanish medium-class population since we had a similar number of subjects from public and private schools (6-year-olds: 202 public and 136 private; 10-year-olds: 76 public and 64 private) who all lived in medium-big cities comprised in the first quartile of the National Economic Ranking[36] (which assesses the income per person in populations with > 1.000 inhabitants).

[36] This data has been obtained from http://elpais.com/especiales/2016/renta-per-capita-municipios-espana/tabla.html, based on the reports that the Spanish Statistical Institute published in July 2016. Those data comprised the 2.964 cities of >1.000 inhabitants in Spain (last accessed 2017 January 16).

Table 15.
Sample distribution by condition, role, age and sex.

		Fair				Hyper-fair				Hyper-unfair				No choice			
		Proposer		Responder		Proposer		Responder		Proposer		Responder		Proposer		Responder	
		Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
6 years	N (%)	25 (58%)	18 (42%)	28 (65%)	15 (35%)	25 (58%)	18 (42%)	18 (42%)	25 (58%)	17 (41%)	24 (59%)	22 (54%)	19 (46%)	21 (50%)	21 (50%)	19 (45%)	23 (55%)
	p-value	0.36		0.07		0.36		0.36		0.35		0.76		1		0.64	
	Totals N (%)	43 (25%)				43 (25%)				41 (24%)				42 (25%)			
10 years	N (%)	13 (68%)	6 (32%)	8 (42%)	11 (58%)	9 (50%)	9 (50%)	12 (67%)	6 (33%)	10 (59%)	7 (41%)	11 (65%)	6 (35%)	7 (44%)	9 (56%)	7 (44%)	9 (56%)
	p-value	0.17		0.65		1		0.24		0.63		0.33		0.8		0.8	
	Totals N (%)	19 (27%)				18 (26%)				17 (24%)				16 (23%)			

Table 16.
Sociodemographic information of the cities where we conducted the experiments.

City	Inhabitants (2016)	Income per person	National Economic Ranking (from 1 to 2.964)
Badalona	215.634	23.860€	512
Barcelona	1.608.746	32.954€	47
Lloret de Mar	37.042	21.720€	850
Rubí	75.167	24.730€	396

Informed consent

We obtained informed consent from the school authorities and parents. The informed consent contained the following information about the study (translated from Spanish, capital letters in the original): "During these tasks, your son/daughter will share stickers with another peer. The University of Barcelona aims to study how children react in front of potentially unfair options and to what extent they are generous between each other (…) IT IS VERY IMPORTANT, if you sign this informed consent, NOT TO MENTION THE PURPOSE OF THIS STUDY to your son/daughter, since this could influence in his/her natural behavior. Do not worry about his/her responses. Since this is just a descriptive and anonymous study, no judgments about him/her will be made. Help us to know better how children are".

Study 2: materials

Apparatus

We built an apparatus that looked like a theatre curtain. More specifically, we used a wooden structure composed of two vertical poles (66 cm) joined on the upper side by a horizontal metal bar (81 cm). An opaque curtain hung from the metal bar to the surface of the table where the apparatus stood up, covering all the space of the structure except for a 50 cm x 12.5 cm window on the lower side of the curtain, which could be used for transferring objects on the table from one side of the curtain to the other. Depending on the height of the chairs where the participants sat down, a black piece of fabric was sometimes added to that window to reduce the size of the transfer area so that players could not get visual clues about the identity of the partner (i.e. skin color, bracelets) but making sure that they still had full visual access to the whole set of options. The curtain helped us to keep the game anonymous. Each side of the curtain had a different color to differentiate the role of each player—blue for proposer and white for responder (see Figure 14).

Rewards

With regard to the rewards, we avoided using food or money in compliance with the schools' policies. The reluctance to use money with children in bargaining games is also common in other countries, as reported in previous studies (Gummerum & Chu, 2014; Güroğlu et al., 2009; Harbaugh, Krause, & Vesterlund, 2007; Murnighan & Saxon, 1998). Instead, we used round, yellow, smiley-face foam stickers (diameter = 4 cm). Stickers for 6- and 10-year-old children have also been used in previous studies (e.g., Benenson, Pascoe, & Radmore, 2007; Gummerum & Chu, 2014; Liu et al., 2016). Nevertheless, we made sure that the children liked the stickers during pilot testing, where we gave them the opportunity to choose among different types (which varied in color, size, and shape). All of the children consistently selected the smiley-face sticker and were excited when they obtained such stickers. Each distribution of rewards (see Conditions section below) was shown on a different plastic tray (35x22x7.5cm). The base of each plastic tray was painted in two colors: half blue and half white. The stickers for the proposer were laid out on the blue part, and the stickers for the responder were arranged on the white part. Both trays were placed on the table, under the transfer area of the curtain, so that the color on the tray matched the side of the curtain with the same color. The use of different colors allowed the participants to easily

identify their role and the quantities that they and their partner could earn in each distribution, as in similar experiments (Güroğlu et al., 2009). Instead of using their hands, every player had a 22-cm-long plastic stick to push the tray through the transfer area so that no visual clues were provided (e.g., skin color, scars) (see Figure 14).

Figure 14. Apparatus as presented during the test: proposer's side (upper-left) and responder's side (upper-right). Stick that children used to transfer trays and smiley face sticker used as reward (bottom-left). Top view of one condition (8/2 and 5/5) displayed on the trays (bottom-right) from the responder's perspective.

Study 2: conditions

Conditions

We had four different conditions composed of two distributions of stickers each, as in Fehr et al's study (2003). One distribution was the default option and remained the same across all conditions (i.e. 8/2). This default option was pitted against another distribution in each condition: Fair (8/2 and 5/5), Hyperfair (8/2 and 2/8), No Choice (8/2 and 8/2), and Hyperunfair (8/2 and 10/0).

As it can be seen, the conditions varied in the level of unfairness from the responder's perspective. Fair condition confronted 2 vs. 5 stickers, namely an unequal distribution versus a half-half payoff. In order to be generous with his partner, the proposer must incur a cost of 3 stickers (8 vs. 5). Hyperfair condition confronted 2 vs. 8 stickers, namely an unequal distribution versus a "super-generous" distribution. We use the adjective super-generous to reflect that in case the proposer offered this option he would keep fewer stickers than the responder incurring a cost of 6 stickers (8 vs. 2) (hence the name "hyperfair"). No Choice condition confronted the same two unequal distributions. The proposer could not be intentionally selfish since the distributions were preassigned. Hyperunfair condition confronted 2 vs. 0 stickers, namely an unequal distribution with an even more unequal payoff (hence the name "hyperunfair"). In order to be generous, the proposer must incur a cost of 2 stickers (8 vs. 10).

Both the proposer and the responder knew that the options were preassigned so that the proposer only had power of decision over which option to offer. As the conditions were preconfigured, this would imply that the proposer's intentions would be transparent in all conditions except for No Choice: if the proposer chose the unequal 8/2 option in Fair condition, this meant that he purposefully wanted to incur no cost when there was a more generous option available, however if he chose the same unequal 8/2 option in No choice, no intention could be inferred since he has no different option to offer. Indeed, from the proposer's perspective, the rewards he or she was willing to lose to make an acceptable offer for the partner gave an idea of the proposer's capacity to foresee others' preferences and to incur costs to obtain gains (strategic behavior). For example, in the Hyperunfair condition, the proposer should choose 10 to obtain the largest payoff, but this option allocated 0 for the responder. If the proposer was able to foresee a rejection of this option, it would be more strategic to choose 8 instead because it allocated at least

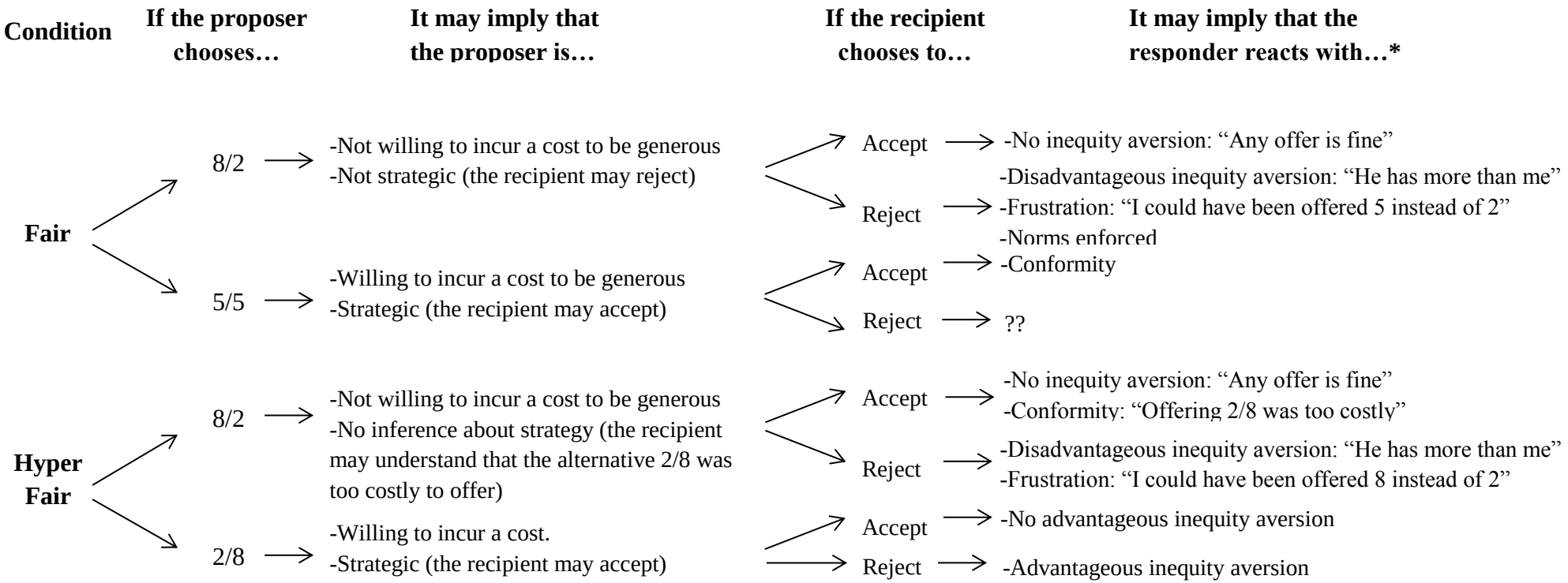
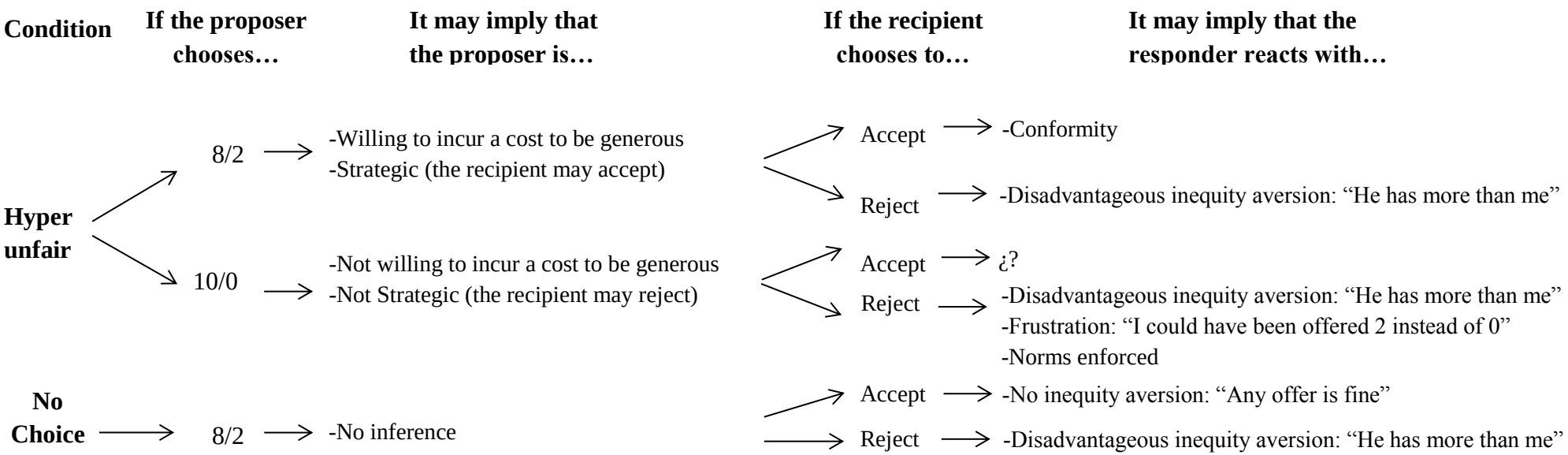


Figure 15. Complete outline of the different responses and interpretations in each condition of the MUG for both players. The scheme continues on next page. * We ran this experiment looking for functional spite, which was defined as costly rejections (for the punisher and for the partner) in one-shot games. In the Introduction of this dissertation I discussed some of the motives that rejections may have (frustration, inequity aversion, seeing norms enforced or making the partner's suffer). I kept the general idea that all rejections in all the conditions can imply the latter motive, but I specified in each condition which of the others could be also present because they depended on the alternative option.



(continuation from previous page)

Figure 15. Complete outline of the different responses and interpretations in each condition of the MUG for both players.

two rewards to the partner. Even though we were especially interested in the presence of functional spite (i.e. rejections), paying attention to the proposers' offer was also important. In case that some age group had a tendency to reject a particular offer, it would be interesting to see whether the same age proposers had a tendency to offer it. Because if they had the tendency to offer what was usually rejected by their peers, it would mean that at that precise age children are able to experience IA or functional spite but they are not able to anticipate the partner's preferences allowed to have an idea about the Similarly, the reactions that the responders had (accept/reject) also allowed us to infer their thoughts about fairness.

Concretely, non-zero rejections gave an idea of the subjective thresholds for what should be acceptable in social comparisons (inequity aversion); rejections of 8/2 offers could be triggered by disadvantageous inequity aversion (i.e. the proposer is getting more than the responder) or by frustration in those conditions when there was a more generous option available (i.e. the proposer could have offered me more); rejections of the alternative option in the Hyperfair condition (2/8) informed about the existence of advantageous inequity aversion (i.e. the responder is getting more than the responder). Finally, one can see whether the responder is sensitive to the proposer's intention regardless of outcome-based considerations. For instance, although an 8/2 offer from the recipient's perspective can be considered unfair from an outcome-based interpretation, it might turn into a fair offer from an intention-based look when the other alternative is 10/0. In other words, the responder could reject an 8/2 offer in Fair condition but accept it in Hyperunfair condition, probably because in the latter scenario it was the more generous available offer. In Figure 15 there is a description of all the different responses that both players could perform in each condition plus their tentative interpretations.

Study 2: procedure

Procedure

In the Introduction we referred to some of the methodological shortcomings that previous MUG studies with children had (see Table 13). Here we first describe point by point how we tried to overcome them (direct method, payoff, anonymity, one-shot) and after that we provide the full procedure we followed in each school[37].

Direct method

To our knowledge, so far only one experiment (Wittig et al., 2013) aside from us (Bueno-Guerra et al., 2016) has conducted MUG with children using the direct-method. In that study, the subjects played side by side in full view of each other with an apparatus that delivered the rewards. Our subjects played in the same experimental room one in front of the other (with an opaque curtain in between to avoid recognition) and interacted live and direct with him/her (i.e. saw how the partner offered or accepted/rejected).

Payoffs

In order to avoid conversion rules and make the payoffs clear, we did use a "what you see is what you get" procedure. We used stickers as rewards, and the quantities the participants saw while playing were the actual payoffs they could end up having. We invested 600 € in buying enough stickers to cover all the potential responses.

Anonymity

Due to all influence that anonymity has on responses, we decided to keep interactions anonymous. This, however, was not always easy to address. In order to have direct interactions (direct-response method) which were anonymous, we used an opaque apparatus (see section Materials above) that impeded both players to see each other. I would place the apparatus in a way that when the second child entered in the experimental room, the first child to have entered was hidden behind the curtain. The order of entrance of the children was indeed an important matter to keep anonymity. I conducted the MUG in the experimental rooms I was assigned to at schools, usually close to the children's classrooms. In schools where there was only one participating

[37] We have recently published the following paragraphs in a shorter version in Bueno-Guerra et al, (2016) Do sex and age affect strategic behavior and inequity aversion in children? *JECP*, 150, 285-300. Here I reproduce and expand further the content of the mentioned publication.

class (that happened twice), the difficulty was to turned the game into anonymous if the players would be taken out of the class by me simultaneously. As a solution, I elaborated an ad-hoc sociometric questionnaire and passed it in class the day before the game so that I could pre-assign the dyads avoiding friends or non-friends. Therefore the game was not anonymous but at least we were sure that the players had a neutral relation with each other. In the schools where there were two participating classes I pseudo-randomized the assignation of roles (proposer/responder) so that one child from one class always played with one child of the other (if the number of children was odd, the remaining child would play with a partner that had already participated before, but their results were excluded from analysis). The day of the game I would first enter in one class and take the first player out; then went to the experimental room and asked the children to stay quite behind the apparatus; then quickly went to the second class (sometimes upstairs!) to take the second player out and finally entered with him/her in the experimental room and start the experiment with both. If for some reason the first player showed his/her face to "nose around", I would stop the experiment and return to their classes to start the procedure again with different children. Those subjects were postponed for a different day with a different partner.

Despite the efforts to keep anonymity by building an apparatus and pre-assigning dyads, I also needed to take care of different aspects of the experimental room, with special attention to the second child to enter. For example, I attached a piece of cardboard under the table so that child could not have any visual clue of his/her partner clothes and shoes, which sometimes identified the children (eg. when a child is wearing new shoes for the first time they always tend to proudly announce it so that the rest of the children associate those shoes with him/her) (see Figure 16). In case there was a sunny day and the child's silhouette could be identified on the curtain against the light (eg. a child with a characteristic haircut), I would attach some black cardboards to the windows to make the curtain absolutely opaque. Finally, children used plastic sticks to push the trays through the transfer area, so that no visual physical clues were provided (eg. skin color, bracelets, scars). When the game was finished I randomly asked a quarter of the children from the schools where there were two participating classes who they thought they were playing with, and none of them gave a right answer.

Figure 16. Typical experimental set-up at schools. The elements that provide anonymity are: opaque curtain, black piece of material in the transfer area, cardboard below the table and the use of sticks for transferring the tray (not visible in the photo).

One-shot

The problem with one-sot games is sample size, since many participants are needed to be recruited for being able to run the statistical analysis. All our dyads played only one condition to avoid gaining any experience about potential partner's responses that could affect their decisions (i.e. one given proposer could decide not to offer 8/2 in the second condition presented if he had experienced a rejection of an 8/2 offer in the condition presented immediately before).

In sum, we provided (1) a real scenario by using the *direct-response method* instead of the strategy method, so that subjects made decisions with both the rewards and their partner present in the same room, as in real markets; (2) we used *clear payoffs* through a "what you is what you get" rule, namely the distribution of rewards that subjects played were the actual payoffs they could get, without any type of conversion rule afterwards; (3) we imposed *anonymity* by pre-assigning the participants of the dyads plus controlling the entrance of the participants in the experimental room; by playing MUG

with an opaque apparatus and by adapting the experimental room to the potential clue getters so that subjects could not obtain any information about their partner other than the fact that he/she was the same age and (4) we avoided any learning effect by letting children play only one role and one condition (*truly one-shot game*).

Full Procedure

Typically, the experimenter visited the school, introduced herself to all the students in the class, explained how the MUG worked and answered any questions that arose, until the game was fully understood (this usually took around 15 minutes). To describe the game, she placed all the material (apparatus, plastic trays, stickers and two sticks) on a table, asked the children to surround it, explained the rules, chose some volunteers to demonstrate the game, and asked them the meaning and outcome of every possible action (offer, accept and reject). This ensured that the children knew that they could decide to either accept or reject the offer. In addition, to prevent these examples from contaminating the children's choices, we used different stickers and quantities from those in the test, as in previous studies (Sutter, 2007).

Subsequently, participants were told that the game was one-shot and that players were to be the same age but anonymous, meaning that no information related to their play mate would be provided (such as their name or ethnicity). Each pair of children played only one condition (Fair, Hyperfair, No choice or Hyperunfair), and every subject played only one role (proposer or responder).

We tested children outside of the classroom in a quiet room at their school. The children entered in the room as specified in the Anonymity point (see above). To ensure that the children fully understood the task after the explanation of the game in class, we did the following: first, before entering the experiment room, every participant was told his or her role (proposer or responder) and was also asked two control questions (he was shown a piece of cardboard with an example of the game and had to answer how many stickers each participant would earn in the case of rejection or acceptance of the offer). None of the children made a mistake. Second, during the game, the experimenter informed the participants about the available options, and gave instructions to follow in the case of acceptance or rejection.

The task took approximately 5 minutes per pair to complete. We videotaped the task with a SONY HDR-CX210 video camera. Participants noticed its presence, but all appeared comfortable and it did not seem to influence their choices. The game started when both participants were in the experimental room, facing the apparatus. The experimenter "read" the options in a loud voice (e.g. "*Here you have two trays. In this one, there are X stickers for you* (pointing to the proposer) *and X stickers for you* (pointing to the responder). *In that one…"*). The proposer had 5 seconds to push one of the trays forward with the stick. Then, the experimenter "read" the offer that was chosen and the actions that could be taken (e.g. "*He is telling you: X stickers for you, X stickers for me. If you like the deal, take the stickers and push the tray back with your stick. If you do not like it, just push the tray back with your stick. Choose whatever you want*") without providing any clues about the partner's identity (the English translation uses "He" as the subject of the sentence, but the expression in Spanish and Catalan is ambiguous). Five seconds were counted before the responder decided. To accept the offer, the responder took the stickers on his or her side out of the tray, and pushed it back to the proposer so that he or she would know that the deal had been made, and thus was allowed to take his or her respective stickers too. To reject the offer, the responder simply pushed the tray back to the proposer, so that he or she would know that no deal had been made as the stickers remained untouched, and thus the game was finished with a zero outcome for both players. When the offer was 0 in the Hyperunfair condition (10/0), as rejection could not be guessed by the proposer, the responder was asked to nod or shake her head in silence, meaning yes or no, respectively. Then, the proposer was informed about this by the experimenter (i.e. "*Your partner has taken the stickers. There is a deal, so now you can take yours.*" or "*Your partner has not taken the stickers. There is no deal*"). Once the game had finished, the experimenter asked some of the participants (both proposers and responders) about their decisions, to explore their comprehension of the MUG. The participants did not take their stickers with them after the game, but the quantity was written down and delivered later that week, when all the subjects had completed the game. We did not give the stickers right after the game to prevent learning or copying others' numbers of stickers by children who had not yet participated. All the children who participated in a game in which the deal was rejected, as well as students who did not have parental consent to participate, received two stickers (the minimum offer), in order to compensate for extreme outcome differences in class between classmates.

Study 2:
results

Results[40]

We divided the results into four types of analysis: offers (age and sex), rejections (age and sex). We conducted two types of non-parametric statistical tests with effect size calculations. We carried out Pearson's $\chi 2$ tests or Fisher's exact test (whenever the assumptions for the former test were not met, i.e. when there were very low expected frequencies) to compare the offers and rejection rates across ages and sexes. Cramer's V was computed to assess the effect size whenever necessary. Finally, when we analyzed the offers, we excluded the No Choice condition (8/2 and 8/2) from the analysis, given that both offers were the same and no information regarding the proposer's preferences could be extracted.

Offers

The frequencies of the 8/2 choices by condition and age are shown in Table 17. The first analysis we did was to compare them across all conditions within each group of age. We did not find any significant results except for the older girls (see Table 18).

Table 17.
Frequencies of choices (offers and rejections) across conditions by age.

		Fair		Hyper-fair		Hyper-unfair		No choice	
		Off*	**Rej****	**Off**	**Rej**	**Off**	**Rej**	**Off**	**Rej**
8/2 Option	**6y**	20 46.5%	8 40%	30 69.8%	7 23.3%	24 58.5%	3 12.5%	42 (-)	4 9.5%
	10y	3 15.8%	2 66.7%	13 72.2%	7 53.8%	16 94.1%	7 43.8%	16 (-)	6 37.5%
Alternative Option	**6y**	23 53.5%	1 4.3%	13 30.2%	3 23.1%	17 41.5%	14 82.4%	-	-
	10y	16 84.2%	0 0%	5 27.8%	0 0%	1 5.9%	1 100%	-	-

* N (%): Counts and percentages over total of offers within experimental conditions.
** N (%): Counts and percentages over offers of the specified option within experimental conditions.

Age. Figure 17 presents the percentage of children who offered non-8/2 options as a function of age and condition. Ten-year-old children were more likely to offer fair offers in the Fair (n=62, $\chi 2$=5.33, p=0.021; V=0.26) and Hyperunfair conditions (n=58, $\chi 2$=7.11, p=0.008; V=0.31) than the younger subjects. In contrast, there were no differences between age groups in Hyperfair: around 30% of both younger and older children offered option 2/8.

[40] This section is reproduced from the article published (Bueno-Guerra et al., 2016)

Table 18.
Comparison of 8/2 offers between conditions within each group of age separated by sex.

					95% CI	
			Difference	p value*	Lower bound	Upper bound
6 years	Girls	Fair-Hyperfair	0.16	0.95	-0.15	0.47
		Fair-Hyperunfair	0.21	0.95	-0.14	0.56
		Hyperfair-Hyperunfair	0.05	1	-0.3	039
	Boys	Fair-Hyperfair	0.33	0.23	-0.01	0.68
		Fair-Hyperunfair	0.04	1	-0.31	0.39
		Hyperfair-Hyperunfair	-0.29	0.23	-0.6	0.02
10 years	Girls	Fair-Hyperfair	0.78	**<0.01**	0.41	1
		Fair-Hyperunfair	1	**<0.01**	0.91	1
		Hyperfair-Hyperunfair	0.22	0.41	-0.15	0.6
	Boys	Fair-Hyperfair	0.17	1	-0.48	0.81
		Fair-Hyperunfair	0.36	1	-0.27	0.99
		Hyperfair-Hyperunfair	0.19	1	-0.34	0.72

*P values were corrected using Holm's method

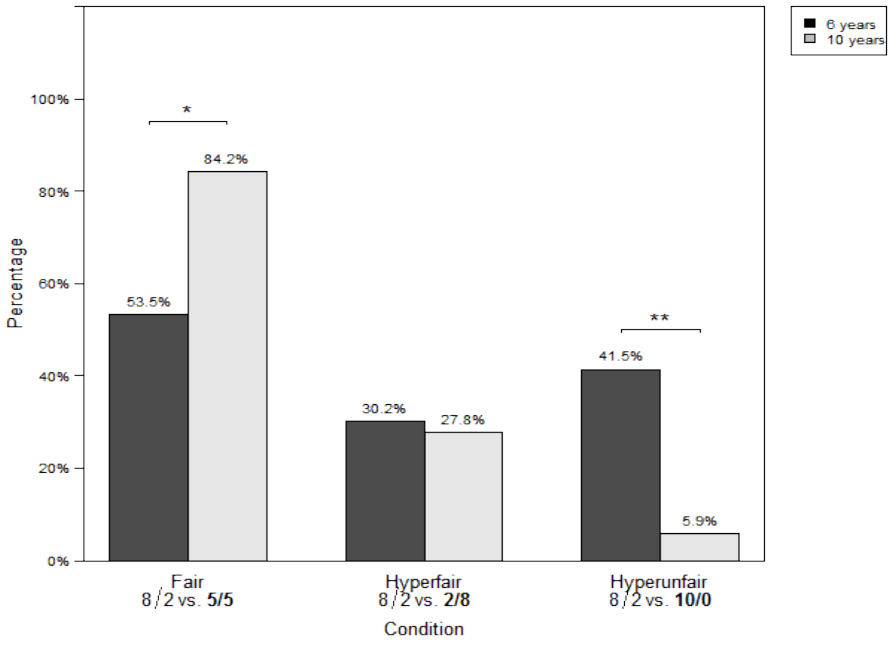


Figure 17. Percentage of alternative offers (other than 8/2, in bold in the graph) as a function of condition and age. * p<.05 and ** p<.01.

Sex. There were no sex differences in the younger children (Fair: n=43; $\chi 2$=0.01, p=.94; V=0.06; Hyperfair: n=43; $\chi 2$=1.71, p=.19; V=0.25; Hyperunfair: n=41; $\chi 2$=0.12, p=.72; V=0.11) (see Figure 18a). Differences appeared with age. At 10 years old, girls significantly increased more strategic behavior by doubling their 5/5 offers in the Fair condition (n=38; $\chi 2$=6.05, p=0.014; V=0.4) and suppressing the 10/0 offer, although the latter change was not significant (n=27, Fisher p=0.057, V=0.32). In contrast, boys maintained their non-strategic behavior, without significant changes according to age. Therefore, in a comparison of both groups, girls behaved in a significantly more egalitarian and strategic way than boys at 10 years of age (n=19, Fisher p=0.04, V=0.48). Remarkably, a hundred per cent of the girls made 5/5 choices in the Fair condition, but only fifty per cent of the boys selected this option (see Figure 18b). Moreover, boys still showed selfish and less strategic behavior in the Hyperunfair condition (i.e. choosing 10/0 instead of 8/2), although the difference was not significant.

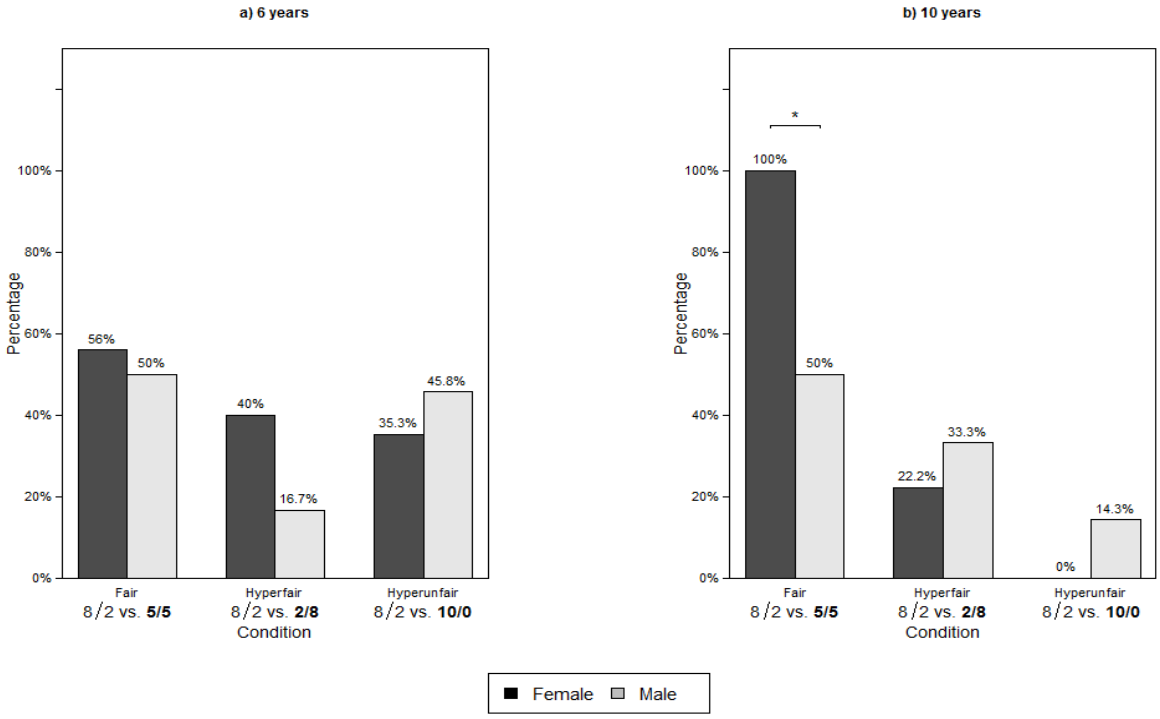


Figure 18. Percentage of alternative offers (other than 8/2) as a function of condition and sex in 6 (Figure 18a) and 10-years-old (Figure 18b) children.

Rejections

The frequencies of the 8/2 rejections by condition and age are shown in Table 17. The first analysis we did was to compare rejections of the default and the alternative options across all conditions within each group of age. We did not find any significant results

except for the alternative option in Hyperfair/Fair against Hyperunfair condition in younger children (see Table 19).

Table 19.
Comparison of 8/2 and alternative rejections between conditions within each group of age.

					95% CI	
			Difference	p value*	Lower bound	Upper bound
8/2 option	6 years	Fair-Hyperfair	0.17	1	-0.14	0.47
		Fair-Hyperunfair	0.28	0.4	-0.02	0.57
		Fair- No choice	0.3	0.08	0.04	0.57
		Hyperfair-Hyperunfair	0.11	1	-0.13	0.35
		Hyperfair-No choice	0.14	0.81	-0.07	0.34
		Hyperunfair-No choice	0.03	1	-0.16	0.22
	10 years	Fair-Hyperfair	0.13	1	-0.6	0.85
		Fair-Hyperunfair	0.23	1	-0.55	1
		Fair- No choice	0.29	1	-0.49	1
		Hyperfair-HyperUnfair	0.1	1	-0.33	0.53
		Hyperfair-No choice	0.16	1	-0.27	0.59
		Hyperunfair-No choice	0.06	1	-0.34	0.46
Alternative options	6 years	Fair-Hyperfair	-0.19	0.24	-0.49	0.12
		Fair-Hyperunfair	-0.78	**<0.01**	-1	-0.53
		Hyperfair-Hyperunfair	-0.59	**0.01**	-0.95	-0.23
	10 years	Fair-Hyperfair	-	-	-	-
		Fair-Hyperunfair	-1	0.11	-1	-0.47
		Hyperfair-Hyperunfair	-1	0.33	-1	-0.4

*P values were corrected using Holm's method.

Age. Regarding 8/2 rejections, overall both age groups had the same decreasing pattern across conditions. As can be seen in Figure 19a, rejection of 8/2 was higher when a more advantageous option was available (Fair and Hyperfair), whereas rejection of 8/2 decreased, but still existed, when it was the best or the only option, but the proposer would still keep more rewards (Hyperunfair and No Choice). However, there were

significant differences in No Choice (n=58, $\chi 2$=4.55, p=.033; V=0.28), in which older subjects were more likely to reject the offer.

Regarding the rejection of alternative options, overall there was no difference between age groups. Only three 6-year-olds rejected 2/8. In general, both groups showed similar pattern to adults: as it can be seen in Figure 19b, generous offers were rarely rejected (Fair and Hyperfair), whereas 10/0 offers (Hyperunfair) accumulated the highest percentage of rejections.

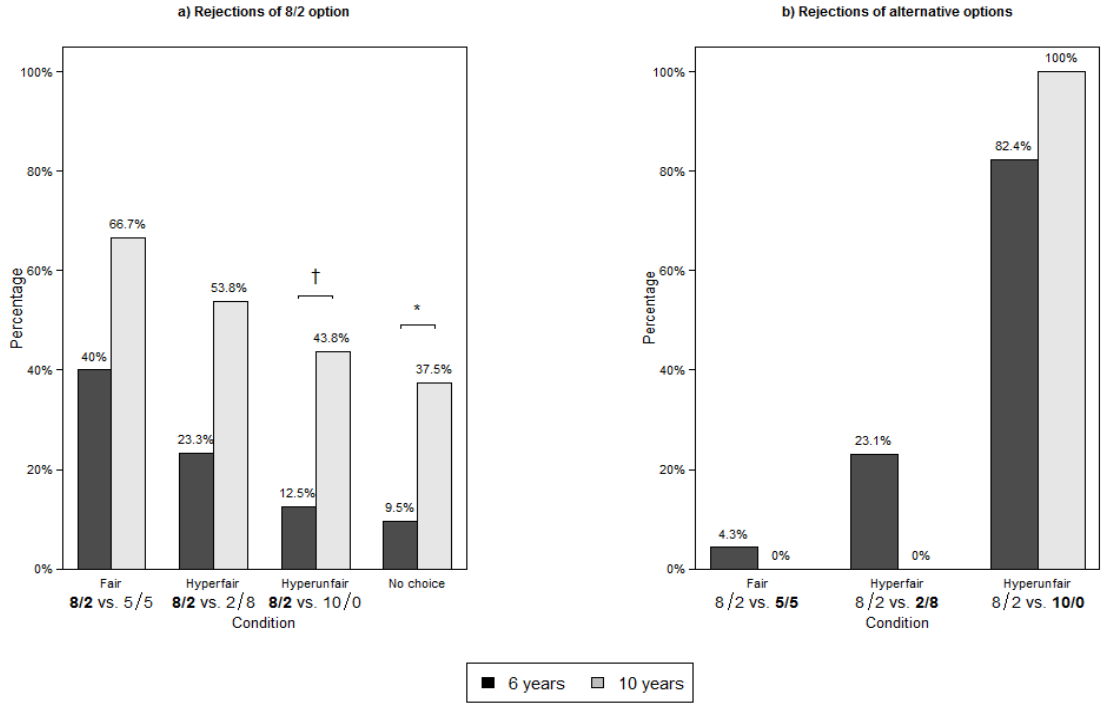


Figure 19. Percentage of 8/2 (left, 19a) and alternative (right, 19b) rejections as a function of condition and age. Ł p<.10 and *p<.05. On the right, rejections of 5/5 and 2/8 offers at 6 years old corresponded to one and three children, respectively.

We also wanted to get an idea about punishment and reinforcement of proposer's intentions. To do this, we analyzed the difference between the percentages of rejection per condition. If responders punish unfair deals based in outcomes, they will reject the default option in Fair and Hyperfair conditions because there is a higher alternative option available or because the partner earns more than them. Statistically, the difference in rejection between options should be significant in both cases (default option mostly rejected; alternative option mostly accepted). Thus we should find significantly more 8/2 rejections than 5/5 or 2/8 rejections. As expected, there were significant differences in rejection between the options in the Fair condition, both at 6

years (n=43, $\chi 2$=6.20, p=.013; V=0.38) and 10 years of age (n=19, Fisher's p=0.018; V=0.56). Both age groups significantly punished non-strategic proposers (those who offered 8/2 when they could have offered a better deal for the responders). The same should be found for the Hyperfair condition, but we decided to exclude it from the analysis because outcomes and intentions might be mixed (i.e. recipients could understand that 2/8 was too costly to offer for the proposer and somewhat forgot him in advanced).

On the other hand, if children acted under the consideration of intentions, we should find significant differences between the rejections of the default condition when the proposer had an equitable alternative available (Fair) and when he had no better option (No Choice). When we looked for it, there were no significant differences despite the absolute difference of around 30% between both conditions in both age groups (see Tables 17 and 19), possibly due to the smaller sample size in those particular subsamples. Another way to investigate whether children take the proposer's intentions into account would be comparing the rejections within the condition Hyperunfair. If responders reinforce intentions, they will accept an offer even though the proposer keeps more than them, as long as the other option available was unfair (accepting 8/2 in Hyperunfair). Statistically, there should be significant differences in rejection between options (8/2 mostly accepted; 10/0 mostly rejected). This was the case at 6 years (n=41, $\chi 2$=20.057, p<.001; V=0.7), but not at 10 years of age (n=17, Fisher's p = 0.47, V≃0), because older subjects rejected 8/2 meaningfully (43.8%). Moreover, 10-year-olds showed a similar 8/2 rejection in No Choice (37.5%), when the proposer could not have any intention to offer unfairly. Therefore, it seems that 6-year-olds took the proposer's intentions into account whereas inequity aversion might have masked this in older children.

Sex. No significant differences were found between boys and girls in rejections (Fair: n=23, Fisher's p = 1.0, V≃0; Hyperfair: n=13, Fisher's p = 0.56, V=0.04; Hyperunfair: n=17, Fisher's p = 0.58, V=0.03; No choice: n=42, $\chi 2$=1.07, p=.74; V=0.05).

Note: A random sample of 25% of the videos was coded by a second observer, who was blind to the experimental design and question. She coded the proposer's choice (which tray she pushed) and the responder's response (grabbing the stickers or pushing the tray back). Inter-observer reliability for both measures was perfect (κ = 1).

Table 20.
Summary of the results in DG/MUG experiment with chimpanzees.

Concept	Key points
Proposers MUG	Proposers did not anticipate the recipient's responses.
	Girls offered more cooperatively than boys at 10 years of age.
Functional spite	There were rejections even though the game was one-shot.
	Boys and girls rejected seemingly.
	6-year-olds rejected taking intentions into account.
	6-year-olds (some) showed advantageous inequity aversion.
	10-year-olds did not reject taking intentions into account.
	10-year-olds presented high rates of disadvantageous inequity aversion.

Study 2: discussion

Discussion

Our results show that by the end of primary education boys and girls endorse cooperation differently: only 10-year-old girls matched their pattern of proposals with the pattern of rejections found in their peers. With regard to functional spite, younger children seemed to take the proposer's intentions into account whereas older children's pronounced aversion to inequity may have prevented us from detecting it: between groups, older participants rejected the default option significantly more than younger ones in conditions where there was no better option available (No Choice and Hyperunfair). Interestingly, this made the young's pattern of rejections more similar to the pattern found in adults (Falk et al., 2003), suggesting a concrete period of high disadvantageous IA (and maybe of high functional spite?) and social comparison during late childhood.

Even though responders frequently punished partners who did not choose the equitable option, and most of them also rejected the 10/0 option, as proposers, they did not consistently offer the fairest option available in these conditions considering their partner's outcomes. This is also what Harbaugh and his colleagues (2007) found in a learning UG experiment. In their study, children from 8 to 18 years played five rounds of UG. After each round, the experimenters showed to the children the responder's behavior of every proposal (i.e. children could see which amounts had been rejected and which ones accepted). The authors found that proposers did not change their offers to match their partners' response: half of them chose 8/2 in Fair and almost 42% offered a 10/0 option. This is very relevant for the study of punishment: because proposers are not able to change his behavior nor to predict which is the best offer to provide, functional punishment does not work in iterated games (Harbaugh et al, 2007) and functional spite exists in one-shot games (our study).

There was the exception of older girls, though. Whereas girls increased their prosocial offers (older girls always offered 5/5 and never offered 10/0), boys, in contrast, made similar offers across the two age groups, and showed a preference for keeping more items for themselves (half of them offered 5/5, and some still offered 10/0 when they were older). These sex differences are in line with previous findings (Murnighan & Saxon, 1998) and might indicate that, at least at 10 years of age, sex would be a discriminative factor in the development of prosocial considerations, so that girls are

readier than boys to form cooperation with other peers. The existence of no-cooperative responses (i.e. selfish proposals) do not necessarily imply that children do not *know* what is fair for others, but rather they do not *act* according to that knowledge (Damon, 1977). That was the case in a recent study (Smith et al., 2013) where 3- to 8-year-old children were able to *state* that they and others should share equally but did not actually *carry out* that behavior until 7 or 8 years of age. The reason for this inconsistency, however, is still to be determined. Proposers might be willing to keep the most for themselves, in line with previous studies of resource sharing (Overgaauw, Güroğlu, & Crone, 2012; Rochat et al., 2009) and only occasionally, but not consistently, take their partners' outcomes into account.

Whether cooperative proposals happened might have also been determined by our methodology. If we compare our results with the evidence with children playing MUG so far (see Table 21 and Figure 20 below), the progression towards prosocial offers over age is not that clear. For example, it can be seen that children of similar ages made proposals three times lower in two different studies (see Figure 21 below). Hence, the oscillation of the graph across child development seems to indicate the existence of certain turning points. By contrast, if we take the results of every study separately and focus in the difference of proposals made in Fair condition between the groups of age they considered, it yields a different picture. In all of them there is a remarkable tendency towards prosociality with age (see Figure 22 below). The relevance of this consideration lies in how developmental conclusions might vary according to the consistency of the methodology employed.

We also looked at functional spite. Our results are in line with the robust body of evidence that shows a strong tendency to punish unequal offers when there is an equitable alternative available, since both age groups significantly punished proposers who did not offer the 5/5 option (functional spite triggered by frustration or to enforce norms). In addition, as previous studies (Güroğlu et al, 2009 and Sutter, 2007), we have not found a significant difference between the rejection of the default option across conditions within each group of age (see Table 22 and Figure 23 in below). This could be interpreted as if the responders were applying functional spite without taking the proposer's intentions into account. However, we have reported an absolute difference of around 30% between the 8/2 rejections in Fair and No Choice conditions in every age group that might have not been significant due to small sample size after subsampling.

Table 21.

Percentages of choices of the default option across conditions in different MUG studies.

	5[a]	6[b]	[7-10][c]	9[d]	10[b]	[11-15][c]	12[d]	15[d]
Fair	56	46.5	15	41	16	5	30	15
Hyperfair	63	70	68	69	72	62	85	68
Hyperunfair	34	58.5	85	60	94	81	75	88

Note: Default options were 8/2 in all experiments except for 3/1 in Wittig et al. 2013. The data corresponds to children and teens (adults have been excluded. meaning ≥18 years old). Decimals have been rounded (>5 increased in one unit). In case the percentages did not appear literally in the text they have been visually extracted from the graphs, therefore it can slightly vary from the actual results. The letters indicate the study where the data come from (we have excluded Gummerum & Chu (2014) because they did not analyze proposer's choices):

[a] Wittig, Jensen and Tomasello, 2013

[b] Our study, 2016

[c] Sutter, 2007

[d] Güroğlu, van den Bos and Crone, 2009

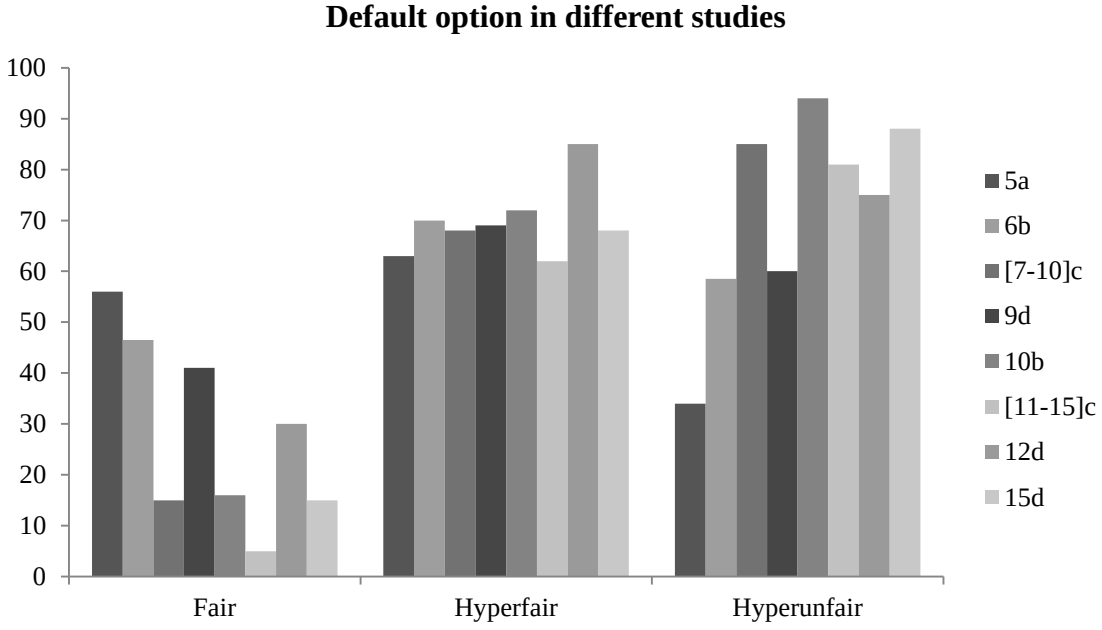


Figure 20. Percentages of choices of the default option across conditions (8/2 in all except for 3/1 in Wittig et al. 2013) from preschool to teens (adults have been excluded. meaning ≥18 years old). Decimals have been rounded (>5 increased in one unit). In case the percentages did not appear literally in the text they have been visually extracted from the graphs, therefore it can slightly vary from the actual results. The letters indicate the study where the data come from (we have excluded Gummerum & Chu (2014) because they did not analyze proposer's choices):

[a] Wittig, Jensen and Tomasello, 2013

[b] Our study, 2016

[c] Sutter, 2007

[d] Güroğlu, van den Bos and Crone, 2009

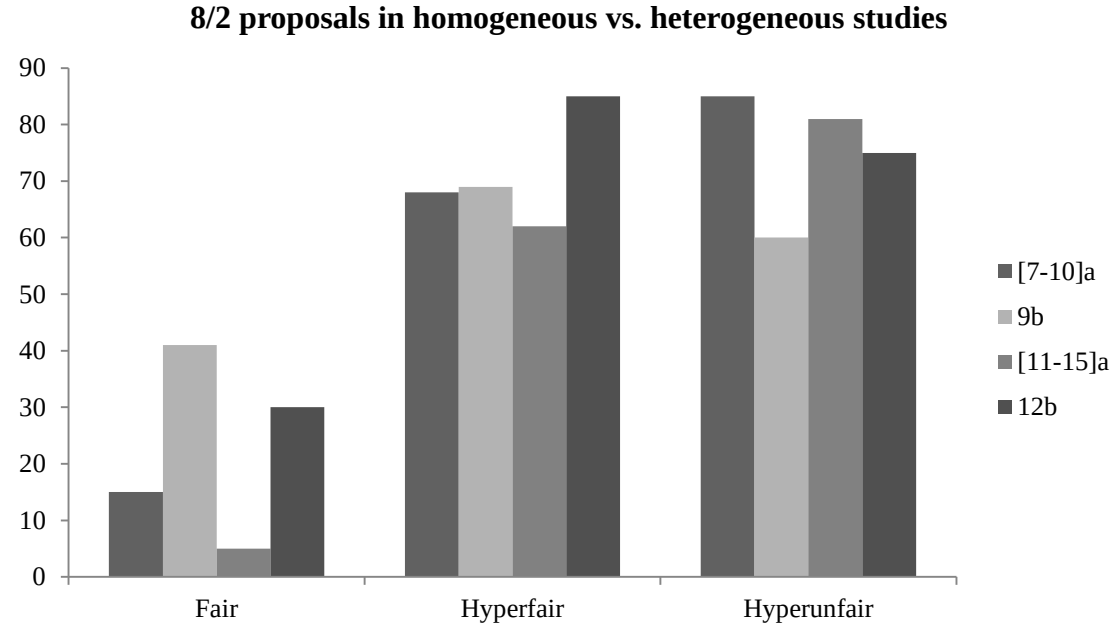


Figure 21. Comparison of percentages of 8/2 choices by proposers in Sutter (2007)'s (letter a in the legend) and Güroğlu et al's (2009) (letter b in the legend) studies. Sutter considered a range of ages of 4 to 5 years (heterogeneous) while Güroğlu et al (2009) considered homogeneous samples (9 and 12 years old). Interestingly, the means of the heterogeneous age ranges of Sutter's study coincide with the means of the homogeneous ages of Güroğlu et al's study. However, the results are not similar but sometimes even the double (ie. Fair condition).

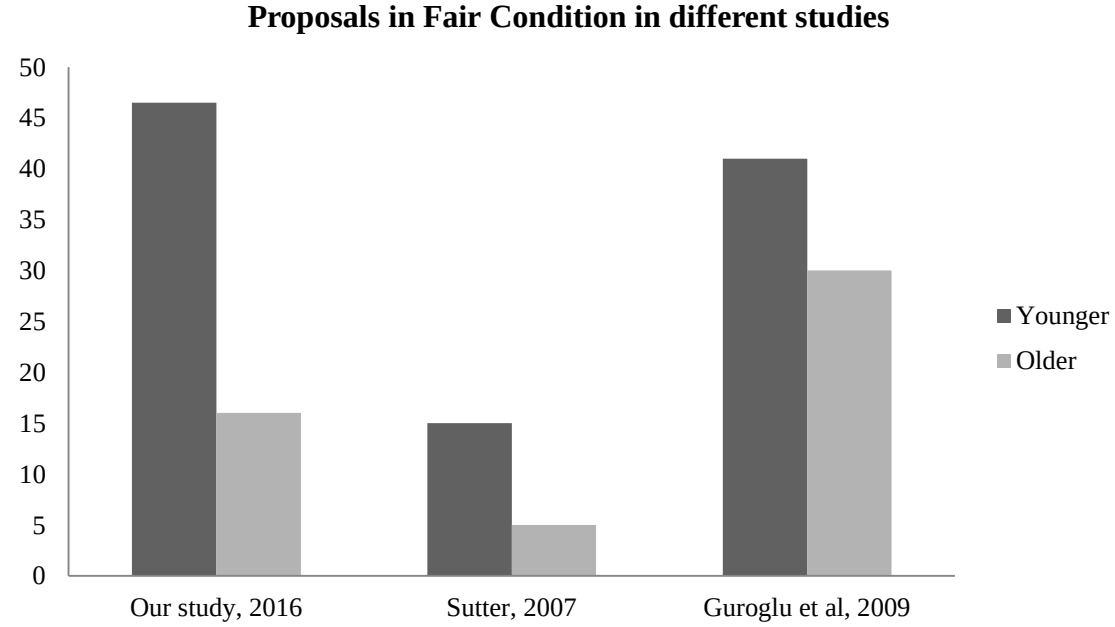


Figure 22. Comparison of percentages of 8/2 choices by proposers in Fair condition across three different studies by age.

In fact, if responders were mainly outcome-based, the more they lose, the more they would reject. Consequently, the default option should yield more rejections in the Hyperfair than the Fair condition, provided that the alternative option in Hyperfair allocated more rewards to the responder. However, that is not the case in all the studies conducted with young children so far (eg. 55% vs. 50% in Gummerum & Chu, 2014; 68% vs 72% in Guroglu et al, 2009; 58% vs. 65% in Sutter 2007; 15% vs 33% in Wittig et al 2013; 23% vs. 40% in ours). Precisely because the differences among the percentages are not large, they contradict the prediction of higher expectation of rejections in Hyperfair. Therefore, it is conceivable that young responders are not only considering outcomes but they might also have some kind of consideration at least about the cost that the proposer would have incurred and that is why they did not react more vigorously. In the case of older children, the general higher percentage of rejections (see below) might have hindered their considerations regarding the proposer's intentions.

Table 22.
Percentages of rejections of the default option across conditions.

	5[a]	**6[b]**	**[7-10][c]**	**9[d]**	**10[b]**	**[11-15][c]**	**12[d]**	**15[d]**
Fair	33	40	65	72	57	70	85	80
Hyperfair	15	23	58	68	54	70	81	75
No Choice	13	9.5	46	65	37.5	47	61	49
Hyperunfair	0	12.5	35	62	44	40	45	45

Note: Default options were 8/2 in all experiments except for 3/1 in Wittig et al. 2013. The data corresponds to children and teens (8/2 in all except for 3/1 in Wittig et al. 2013) from preschool to teens (adults have been excluded. meaning ≥18 years old). Decimals have been rounded (>5 increased in one unit). In case the percentages did not appear literally in the text they have been visually extracted from the graphs, therefore, it can slightly vary from the actual results. The letters indicate the study where the data come from (we have excluded Gummerum & Chu (2014) because they focused on third-party punishment):
[a] Wittig, Jensen and Tomasello, 2013
[b] Our study, 2016
[c] Sutter, 2007
[d] Güroğlu, van den Bos and Crone, 2009

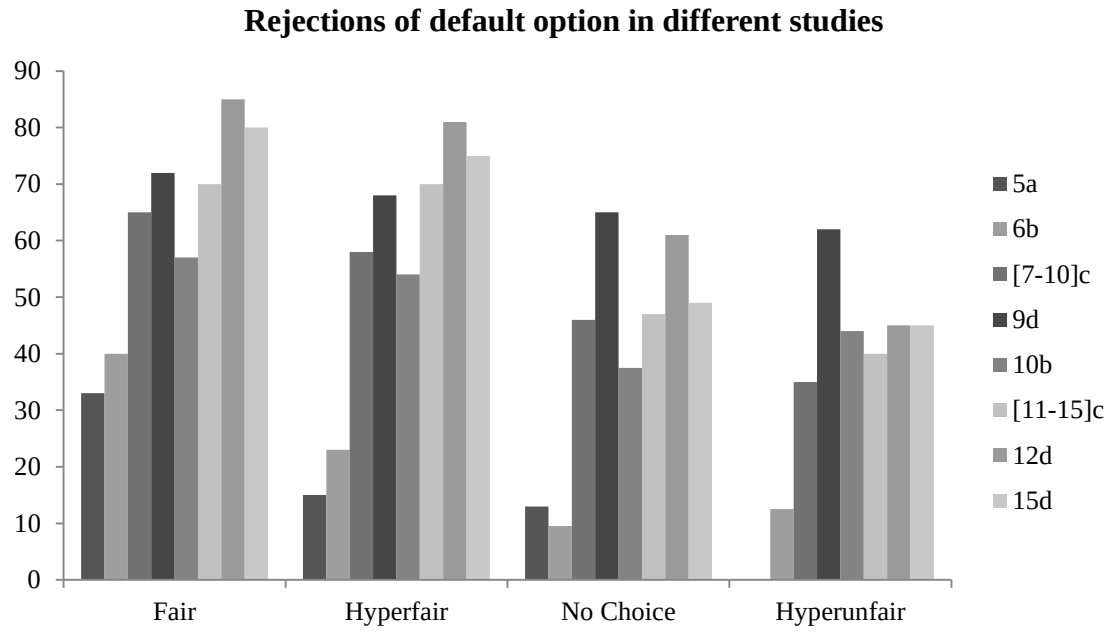
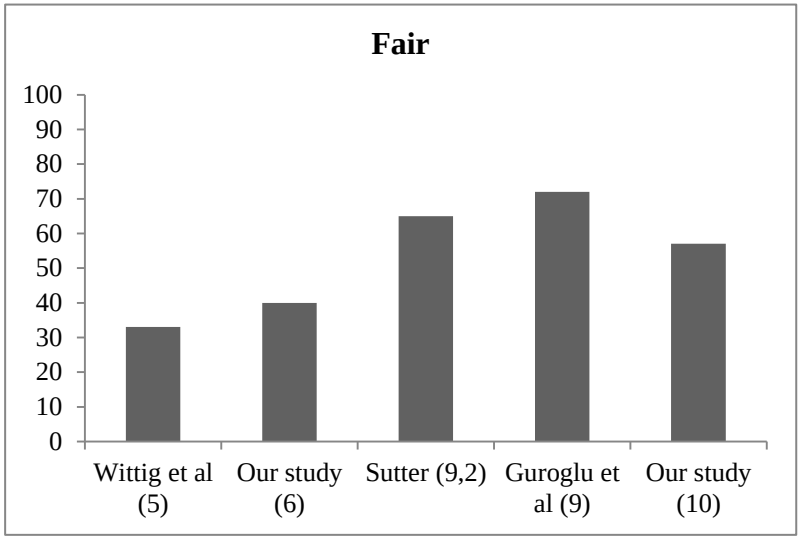
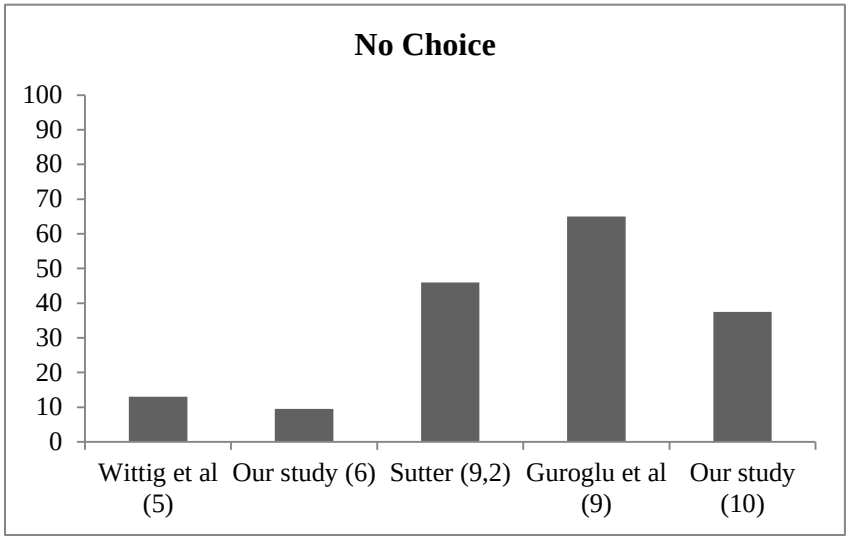


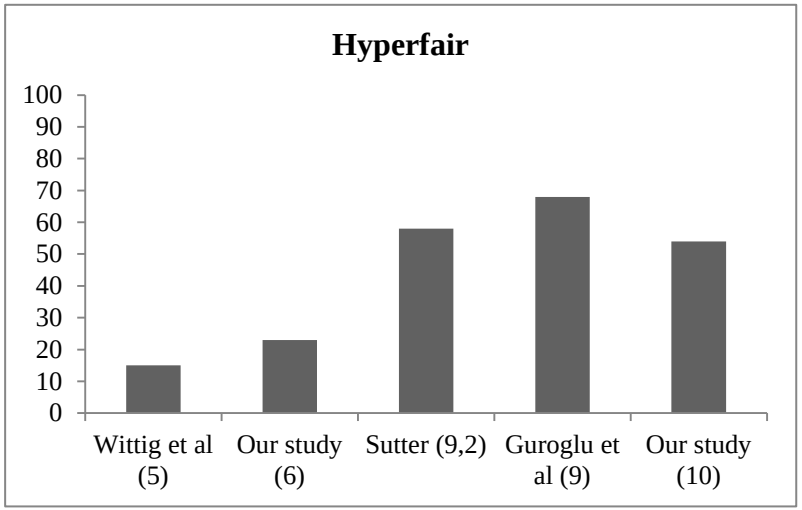
Figure 23. Percentages of rejections of the default option across conditions (8/2 in all except for 3/1 in Wittig et al. 2013) from preschool to teens (adults have been excluded. meaning ≥18 years old). Decimals have been rounded (>5 increased in one unit). In case the percentages did not appear literally in the text they have been visually extracted from the graphs, therefore, it can slightly vary from the actual results. The letters indicate the study where the data come from (we have excluded Gummerum & Chu (2014) because they focused on third-party punishment):
[a] Wittig, Jensen and Tomasello, 2013
[b] Our study, 2016
[c] Sutter, 2007
[d] Güroğlu, van den Bos and Crone, 2009.

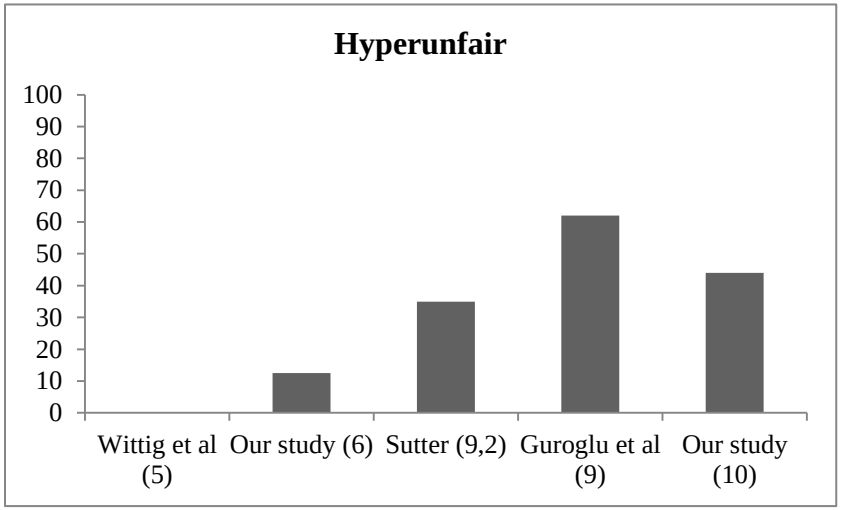
This is the first time that functional spite is reported in 6 years-old within the context of outcome/intentions (and not only outcomes, as in Blake, Piovesan, Montinari, Warneken, & Gino, 2015 or Sally & Hill, 2006). Whereas Wittig and colleagues (2013) show that children by the age of 5 never rejected the default option in Hyperunfair condition, we have shown that 12.5% of the 6 year-olds responders did. This is interesting because it might show that functional spite is differently motivated with age. It might happen that in preschool years children mainly address their look to the outcomes they can get and decide whether rejecting based on that comparison (i.e. frustration), and it is only in absence of a better alternative (No choice) when they turn their look to their partner's outcomes (i.e. disadvantageous IA). Later on, the beginning of school might introduce more peer activities and resource sharing situations, thus promoting a strongest bias for social comparison. That would explain why only just one year later, by 6 years of age, children extend functional spite triggered by disadvantageous IA also to contexts where the offer received was the best option available (see Figure 24 below).

Figure 24. Percentages of rejections of the default option across conditions (8/2 in all except for 3/1 in Wittig et al. 2013) separated by condition.









This is especially prominent at 10 years of age. In fact, the absolute percentages of 8/2 rejections reaches such a high rate then that if we examine Falk and colleagues' data (2003) on adults, the 6-year-olds' pattern of rejection is much more similar than that of the 10-year-olds. This suggests more functional spite that vanishes later in adulthood, probably due to a stronger bias for social comparison in pre-teenagers.

It is worthwhile to refer to those few cases of functional spite triggered by advantageous IA (i.e. rejections of 2/8). In line with the 8% of 5-year-olds that rejected a profitable 1/3 outcome in Wittig and colleagues' study (2013), we have shown that at least three 6 years old children out of thirteen (23%) also rejected the 2/8 offer. It might be argued that these few subjects might have not understood the game, but control questions after the test showed they had. Also, it might be said that individual differences had an influence, but we looked for the potential existence of personality traits while bargaining and found negative results (unpublished). Taken together this evidence may suggest that the emergence of advantageous inequity aversion arises earlier than reported in previous experiments. Previous work had delayed its appearance at about 8 years of age (Blake & McAuliffe, 2011; McAuliffe et al., 2014).

Studies 3 and 4:
Schadenfreude

State of the art

The experience of emotions is a likely proximate cause that motivates costly punishment/spite of antisocial others in humans (de Quervain et al., 2004; Fehr & Gächter, 2002; Hein et al., 2010; Singer et al., 2006). Whereas seeing others suffer typically induces emotional states such as empathy (Singer et al., 2004), which in turn is a powerful motivator for altruistic helping (Batson, 1991; Batson et al., 1981; Hein et al., 2010), this can turned into feelings of pleasure when the suffering victim was previously antisocial or perceived as an outgroup member (Hein et al., 2010; Singer et al., 2006). Such signals of reward have been shown to be critical predictors of a subsequent absence of helping and desire for revenge (Hein et al., 2010; Singer et al., 2006). While there are several studies on this question in young children, nothing is known whether and how such mechanisms might operate in our closest relatives, the chimpanzees. Thus, young infants display an early preference for prosocial compared to antisocial agents (Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007) and prefer those who are antisocial to previously antisocial others (Hamlin, Wynn, Bloom, & Mahajan, 2011). Further, preschoolers have been shown to endorse the misfortune of competitors (Schulz, Rudolph, Tscharaktschiew, & Rudolph, 2013; Shamay-Tsoory, Ahronberg-Kirschenbaum, & Bauminger-Zviely, 2014), to think antisocial others as deserving of punishment (Tisak, 1993) and to punish transgressions of outgroup members more than those of ingroup members (Jordan et al., 2014). While it is known that chimpanzees appear to develop attitudes towards others based on previous pro- and antisocial behaviors (Herrmann et al., 2012; Jensen et al., 2007b; Russell et al., 2008), nothing is known about the phylogenetic origins of the motivation to watch the enactment of revenge.

We studied whether children aged 4-6 years and chimpanzees showed an increased motivation to witness the punishment of an agent that had previously acted antisocially towards them compared to the punishment of an agent that had previously acted prosocially towards them. The pro- or antisocial nature of the agents was operationalized by means of them offering valuable goods to children (i.e. their favorite toys) and chimpanzees (i.e. food). Whereas the prosocial agent would both offer and give the goods to the participant, the antisocial agent would offer the goods first but then withdraw the goods. As indicators of a motivation to witness punishment we used the amount of cost incurred to continue watching the punishment. We operationalized

cost incurred as the expenditure of valuable coins for children and physical energy for chimpanzees. To measure emotional correlates in children we scored their facial expressions (e.g., smiles, frowns) during the punishment of the two agents.

We were also interested in whether, in line with previous work in humans (Singer et al., 2006), there were signs of empathic distress when witnessing prosocial agents being punished. Whereas in children, there is already a wealth of evidence for such basic empathic tendencies when watching others harm themselves (Nancy Eisenberg et al., 1996; Fabes, Eisenberg, & Eisenbud, 1993), as expressed by verbalizations and facial expressions such as frowns (Valiente et al., 2004), for chimpanzees this is virtually unknown. One key indicator of empathic distress is whether individuals have a motivation to escape the distressing situation (Batson et al., 1981). We were therefore interested in whether chimpanzees would incur costs to escape from a situation when observing the punishment of the prosocial agent. For chimpanzees, we also used their vocalizations of arousal (here defined as a compound of distress and display vocalizations, see Procedure and Results for chimpanzees below for more details) during the initial punishment as indicators of emotional arousal. The vocalizations were categorized according to their acoustic and temporal properties28 and grouped according to the call categories suggested by Goodall (Goodall, 1986).

We administered a cross-species forced-choice behavioral paradigm, assessing whether chimpanzees and children incur costs to continue watching the punishment of agents depending on whether these had been pro- or antisocial in a directly experienced previous interaction with them (Study 3). To explore the phylogeny of third-party punishment we also tested chimpanzees in a paradigm (Study 4) in which the chimpanzees did not directly experience but merely witnessed or indirectly experienced, how the prosocial and the antisocial agents interacted with another chimpanzee (stooge). In any other respects, the direct and the indirect tasks were identical. The punishment procedure for all the studies entailed a punisher applying physical punishment in the form of hitting each of the two agents (i.e. see Procedure below).

Crucially, after a brief period of witnessing the punishment, this was rendered invisible to subjects (i.e. occurred in another part of the room for chimpanzees / was occluded by a curtain of a puppet theatre for children). Therefore to continue watching the punishment subjects had to incur costs, which for chimpanzees entailed physical effort

by operating a heavy sliding door to get to the invisible part of the room and for children entailed paying coins for the curtain of the puppet theatre to be raised again.

To test for something akin to empathic distress in the chimpanzees, we included another condition of visible punishment of both agents. The aim of this condition was to see if the punishment of the prosocial agent elicited escape behavior (by operating the heavy sliding door and moving into another part of the room without visual access to the punishment of the agent) a behavior that has been taken as an indicator of empathic distress (Batson et al., 1981). Thus, for the chimpanzees (Studies 3 and 4) this resulted in a 2x2 factorial design with factor prosociality (prosocial/antisocial) and visibility (visible/invisible) and one trial for each condition. For children (Study 3) this resulted in a design with one factor of prosociality (prosocial/antisocial) and with 4 trials for each condition.

To assess whether the prosocial/antisocial exposure procedure had been effective, we assessed the subjects' preference for the prosocial and antisocial agent upon completion of the tasks. For children this was tested by means of explicit preference questions referring to niceness and willingness to share and whom they would play with (see corresponding Procedure below). For the directly (Study 3) and indirectly (Study 4) experienced pro- and antisocial experiments in chimpanzees, this was tested by allowing the chimpanzees to beg for food from the two agents to assess whether they showed a preference for one of them (see corresponding Procedure below). We predicted that both, chimpanzees and children in the direct interaction (Study 3) would be more motivated to watch the punishment of the antisocial compared to the prosocial agent. For the chimpanzees this should entail more subjects incurring physical costs to operate the heavy sliding door during the antisocial compared to the prosocial agent's punishment when this was no longer visible (i.e. antisocial invisible vs prosocial invisible conditions). For the children this should entail subjects incurring greater monetary costs to watch the punishment of the antisocial compared to the prosocial agent. For the children we also predicted differential signs of emotions during the initial punishment of the two agents. Thus, we expected greater number and duration of smiles during the antisocial agent's compared to the prosocial agent's punishment. Happy/positive emotions in chimpanzees are very rarely observed, except in playful activities in which the ape being physically touched (tickled/chased) performs play panting vocalizations (laughter-like) (Matsusaka, 2004). We thus did not have any

specific prediction about the occurrence of such positive vocalizations in chimpanzees as an indicator of joy during the punishment of the antisocial agent.

We also predicted that children and chimpanzees would show signs of empathic distress in response to the punishment of the prosocial compared to the antisocial agent. For the chimpanzees this should entail more subjects incurring physical costs to operate the heavy sliding door to escape during the prosocial compared to the antisocial agent's punishment when this continues to be visible (i.e. antisocial visible vs prosocial visible conditions). For the children we predicted signs of empathic distress as indicated by increased frowns during the punishment of the prosocial agent.

We also implemented a study in which chimpanzees witnessed an interaction between pro- and antisocial agents and another chimpanzee If chimpanzees preferentially watched the punishment of antisocial agents regardless of their involvement in the transgression the pattern of results should be the same in both studies (norm-based punishment instead of anger- and revenge-based). In contrast, if involvement aspects played a more important role, we predicted a different pattern of results between Study 3 and 4. In fact, based on previous studies showing that chimpanzees do not punish others who stole food from third parties (Riedl et al., 2012) we predicted that chimpanzees in Study 4, unlike Study 3, would not care to watch or vocalize differentially when others (regardless of whether they were prosocial or antisocial) were being punished. Note that Study 4 differed from Study 3 only in terms of the extent to which the chimpanzee subjects were directly affected by the agents' behavior, while keeping all other aspects of the experimental set-up constant.

It is important to note that our dependent behavioral variable of opening the heavy sliding door for the chimpanzees is always the same throughout all conditions. However, we interpret it differently depending on the condition (i.e., to continue witnessing the punishment when it is invisible or to escape into another room when it is visible). While we tested three age groups of children, we were agnostic to any age-related changes in our variables of interest. We used two-tailed statistics except for the prosocial condition.

Study 3 (children): sample

Children tested

We selected 4-, 5-, and 6-year-old children due to their age-related differences in overcoming inequity aversion (Fehr, Bernhard, & Rockenbach, 2008) and in non-social processes involving self-regulatory mechanisms (Rothbart, Sheese, Rueda, & Posner, 2011). As a result, the 72 children we tested were divided into three age groups: 24 4-year-olds (M= 4.15, age range= 4.04-4.35), 24 5-year-olds (M= 5.04, age range= 4.97-5.4), and 24 6-year-olds (M= 6.17, age range= 5.98-6.33). In each group there were the same number of boys and girls. Seven children were excluded from the analyses due to procedural error or ambiguous behavior. All remaining subjects received all conditions. The children were recruited from a database at the MPI-CBS in Leipzig, Germany. The ethics committee of the University of Leipzig approved the study and the caregivers provided written consent form.

Sociodemography of the city

Leipzig is the largest city (570.087 inhabitants in 2015[41]) in the federal Saxon state in Germany and the tenth most populous city of the country[42]. The Saxon state occupies the 8th position of the German gross domestic product ranking, a central position out of the sixteen states. The kindergarten from which the children were recruited comprised families from medium economic backgrounds that are open to research and often collaborate in developmental studies.

[41] This data comes from: http://www.leipzig.de/news/news/leipzigs-einwohnerzahl-knackt-die-570-000/ (Last accessed 2017 February 3).
[42] This data comes from: https://www.statistik.sachsen.de/html/358.htm (Last accessed 2017 February 3).

Study 3 (children): materials

Apparatus

The apparatus was the same across all the phases. It consists of a wooden miniaturized theatre (approx. 1.20cm x 90cm) with a curtain. The wooden structure allowed an experimenter to hid behind, as a puppet show, and move the toys from there. The punisher was a toy from the Institute (i.e. elephant, lion, giraffe) and the toys that were exchanged (prosocial) or kept (antisocial) belonged to each subject (they were brought to the Institute facilities by the caretakers). We also had a child chair, a table, two boxes and coins (plastic circles).

The whole experiment was conducted in the research facilities of the Max Planck Institute for Human Cognitive and Brain Sciences, which consists of a room with integrated cameras decorated with childish elements.

Study 3 (children): procedure

General procedure

This study tested whether children were willing to incur costs to watch how previous prosocial (i.e. empathy) or antisocial (i.e. indirect revenge) agents were punished by a third-party. In order to do so, we divided the procedure into four phases: (starting) + training + exposure + punishment + preference that were performed both with children and chimpanzees.

Starting

The children came into the lab accompanied by at least one caregiver. The caregivers brought six of their child's favorite toys, as we had asked them by phone, which were used as the toys-characters during the experiment. At the beginning of the study, the children were given an initial endowment of 4 coins. The coins were of no real currency but the children were told that at the end of the study each of the coins could be traded for one sticker. As it can be seen in Figure 25, we used a miniature-sized puppet theater. The child sat down on a chair in front of the theatre. There was a table in between the child and the theatre where children could interact with their coins and the boxes.

Figure 25. Experimental set-up for the Schadenfreude experiment with children.

Training

In this phase, children learnt that they could decide whether the play continued or not by using the coins and boxes they had in front of them. To so do, we performed a hide and seek game with two puppets as part of a theatre play. The children could not see the adults moving the puppets. In the table in front of the child, there were two boxes, one on the right and one on the left. After 20 sec. of the performance, the theatre curtain

closed and the subject was told that if he/she wanted to continue watching the performance he/she would have to place one coin into the box on the right. If he/she did not want to continue watching the performance he/she would have to place one of coin into the box on the left. The child was told that all the coins placed into the right (continue watching) were no longer his/hers, whereas all the coins placed into the left (stop watching) could be exchanged against the stickers at the end of the experiment. Therefore, children need to pay in case they wanted the performance to continue. Depending on what the child did, the curtain re-opened and the performance continued for 10 sec. or the performance finished. At the end of the spectacle, the children were asked 7 control questions making sure that each aspect of the experimental procedure had been fully understood and comprehension scores were calculated. These questions were:

1."*If you have x amount of coins, how many stickers will you receive at the end?*"
Correct answer is x amount of stickers.

2. "*Here are two coins. Who do these coins belong to?*"
Correct answer is these coins belong to the child.

3. "*What should you do with your coin if you want to continue watching the puppet show?*"
Correct answer is to put them in the box on the right.

4. "*What should you do with your coin if you do not want to continue watching the puppet show?*"
Correct answer is to put them in the box on the left.

5. "*What can you do with all the coins lying in the box on the right at the end of the show?*"
Correct answer is nothing, they do not belong to me.

6. "*What can you do with the coins lying in the box on the left at the end of the show?*"
Correct answer is I can trade them in for stickers.

Exposure phase

In this phase the children learnt the moral reputation of two different characters. In order to do that, we had the prosocial condition, in which a prosocial puppet gave toys to the child, and an antisocial condition, in an antisocial puppet took the toys away from the child. Before starting the performance, out-of-sight of the child, the caregiver gave three of the child's favorite toys to the experimenter, who hid them behind the theater curtain.

The show started with one puppet, who introduced himself while holding one of the favorite toys in front of the child: "*Hello* (name of the child), *my name is* (name of the puppet). *How are you doing? Ooohh, great. I am also very good … and do you know what? I would like to play with you. Would you also like to play with me? Look, your daddy/mammy brought me three of your toys. Wow, look here* (name of the child)! *It is your* (name of the toy). *It is very nice and I know that you like* (name of the object). *Therefore, I will give it to you*". Afterwards the puppet leaned towards the child and offered him/her the toy. In the prosocial condition, the prosocial puppet let the child take the toy. In the antisocial condition, the antisocial puppet took it back and said "*No! I have decided to keep it for me!*". This procedure was repeated two more times (each time with a different toy). We counterbalanced the order of presentation of the conditions and the identity of the exchanging puppets (i.e., elephant, lion, giraffe).

Punishment phase

In this phase children saw how a third puppet that they had not seen before ("the punisher") started beating one of the previous puppets, and they could decide whether they continued watching the punishment (paying coins) or not (placing coins in a different box to be traded against stickers). The children watched how the punisher started beating the prosocial and the antisocial puppet separately.

This phase started with the introduction of the third puppet, the punisher, who carried a stick: "*Hello* (name of the child). *I will now hurt the* (name of the puppet)". After an initial punishment, which consisted of a few beatings with the stick to the puppet, the punisher turned again to the child and said: "*I will now pursue the* (name of the puppet) *and continue to hurt it. If you want to continue watching how the puppet is hurt then you will have to put one of your coins into this box*" (while pointing to the box). *If you*

do not want to continue seeing how I hurt the puppet (name of the puppet) *then you should drop one of your coins in your box*" (while pointing to the box). Then, the punisher started beating the puppet up with the stick for 5 sec. (i.e., 5 hits with the stick, rate: 1 Hz). After that time the theater curtain closed and both the punisher and the hurt puppet were out-of-sight of the child. The experimenter asked the children if they wanted to continue watching the punishment. Given that children had received 4 coins, the maximum number of paid punishments was 4. Thus, all subjects received exposure to the first round of punishment and the first question of whether they would like to continue watching or not and then, depending on whether children paid for punishment, they were asked again until they either decided to stop watching or until they had no more coins. The final round was always the pursuit and punishment behind the curtain, thus the child continued hearing the puppet crying for 10 more sec. but without visual access to the punishment.

Preference phase

At the end of the entire Punishment phase (after both the antisocial and the prosocial puppets had been attacked by the punisher), the experimenter showed the two puppets to the child and asked which puppet the child would rather play with, give a sticker to and thought was nicer. From this we obtained a composite score of preference.

Coding

All sessions were videotaped and the following variables were coded from digital files during the exposure phase as well as the punishment phase: 1) behaviors and verbalizations 2) pure smiles, pure frowns and given the potential ambivalence of seeing someone antisocial experiencing punishment, we also coded for smiles occurring jointly with frowns. Two observers coded all the videos using the Interact software. To assess inter-observer reliability, ratings were correlated. Inter-observer reliability was high for answering the questions of the punisher ($r = 0.99$, $p < 0.0001$) as well as for occurrence of smiles, frowns and smiles with frowns during the exposure as well as the punishment phase (all $r > 0.504$, all $p < 0.0001$).

Study 3 (children): results

Results

We first confirmed that children had understood the valence of the agents (prosocial and antisocial). To do so, we analyzed their responses to the six control questions and their responses in the preference phase ("Comprehension"). Next, we analyzed whether children were willing to incur costs to watch those agents being punishment and whether they preferred to watch more one over the other ("Schadenfreude results"). Finally, we analyzed their emotional reaction to the exposure and the punishment phases by analyzing the occurrence of smiles and frowns ("Facial expressions").

Comprehension

Comprehension of the task was good (M = 5.7, St. Dev = 1.04). There were no comprehension differences between 4- and 5-year olds ($F(1.41) = 2.9$, $p > 0.1$) or 5- and 6- year olds ($F(1.41) = 2.9$, $p > 0.1$).

When we tested which of the two agents children preferred, the three age groups displayed a clear preference for the prosocial over the antisocial agent (paired t-test: $t(64) = 4.279$; $P < 0.001$) with no age differences in this preference (one-way ANOVA; $P > 0.607$).

Schadenfreude results

To test for the hypothesis that children would show an increased motivation to observe the punishment of a previously antisocial agent, we compared the number of coins spent on continuing to watch the punishment of the prosocial and the antisocial agents. A Repeated Measures ANOVA with agent as a within-subject and age-group as a between-subject factor, indicated a significant interaction between the factors agent and age-group in how coins were allocated to watch the punishment ($F(2,62) = 3.417$; $P = 0.039$, Figure 26 left). Thus, only 6-year-olds allocated more coins to watch the punishment of the antisocial compared to the prosocial agent ($F(1,20) = 12.246$; $P = 0.002$; for 4- and 5-year olds $p > 0.2$; Figure 26 left).

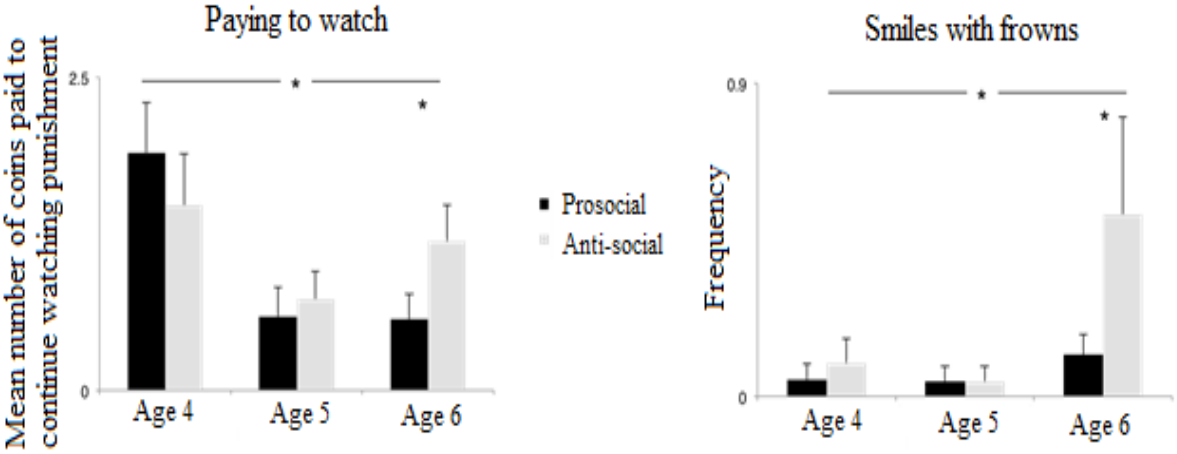


Figure 26. Mean number of coins paid to watch punishment by age and valence of the agent (left). Frequency of smiles in co-occurrence with frowns by age and valence of the agent (right).

Facial expressions

During the exposure phase the facial expressions of the children revealed no differences in number or duration of smiles ($F < 0.73$, $p > 0.39$) and smiles with frowns ($F < 3.1$, $p > 0.1$) nor significant age differences ($F < 1.9$, $p > 0.14$). However, there was a greater number of frowns during the interaction with the antisocial than the social agent ($F(1.62) = 6.632$, $p = 0.012$) and no significant age-difference.

While watching the initial round of punishment, the coding of facial expressions showed significant age-differences in number of smiles co-occurring with frowns depending on which agent was being punished ($F(1,62) = 2.294$; $P = 0.03$, one-tailed; Figure 26 right). Thus, only 6-year-olds showed an increased mixture of positive and negative emotions (facial expressions) while watching the punishment of the antisocial compared to the prosocial agent ($F(1,20) = 3.155$; $P = 0.045$, one-tailed; Figure 26 right). We assessed the number of frowns during the initial round of punishment as an indication of empathic distress in the children at seeing the punishment of the agents. Whereas children frowned for both the prosocial (one-sample t-test: $t(64) = 2.408$; $P = 0.019$) and the antisocial agent (one-sample t-test: $t(64) = 2.644$; $P = 0.010$), this did not differ between the two agents. Frowning during the punishment did not interact further with age ($P > 0.4$).

Studies 3 and 4 (chimpanzees): sample

Chimpanzees tested

This study was conducted across three years (2013-2016), during which some of the initially tested subjects left to a different zoo. That is why not all the subjects participated in both studies. Below there is Table 23 with the information about each subject: age (at the moment of the end of the study, in May 2016), sex, rearing history and the study in which they participated.

Table 23.
Information about the subjects of Studies 3 and 4.

Chimpanzee Group	Subjects	Age (years, months)	Sex	Rearing history	Study
A	Corrie	(12.4)	F	Hand reared	Studies 3 and 4
	Fraukje	(37.1)	F	Hand reared	Studies 3 and 4
	Kofi	(8.4)	M	Mother	Studies 3 and 4
	Lobo	(9.1)	M	Mother	Studies 3 and 4
	Lome	(12.5)	M	Mother	Studies 3 and 4
	Natascha	(33.0)	F	Hand reared	Studies 3 and 4
	Sandra	(20.3)	F	Mother	Studies 3 and 4
	Kara	(8.3)	F	Mother	Study 3
	Riet	(36.8)	F	Hand reared	Studies 3 and 4
	Robert	(38.9)	M	Hand reared	Studies 3 and 4
	Tai	(15.6)	F	Mother	Study 4
	Ulla	(36.3)	F	Hand reared	Study 3
B	Alex	(10.9)	M	Hand reared	Studies 3 and 4
	Alexandra	(12.4)	F	Hand reared	Study 3
	Annette	(12.4)	F	Hand reared	Study 3
	Fifi	(18.7)	F	Mother	Study 3
	Jahaga	(18.11)	F	Mother	Studies 3 and 4
	Trudy	(18.7)	F	Mother	Study 3
	Frederike	(42)	F	Unknown	Study 4
	Daza	(30)	F	Unknown	Study 4

Enclosure and facilities

The chimpanzees belonged to the same zoo that it was already mentioned in Section UG with chimps, so I refer the reader to there for the description of the enclosure, habitat and social structure.

Studies 3 and 4 (chimpanzees): materials

Materials

The Schadenfreude experiment with chimpanzees involved different materials for each phase (see Procedure below). Next, we list them all with their function.

Training

During the training, the chimpanzees learnt how to open a sliding door. Therefore we needed a heavy sliding door (approx. 80cm x 60cm) built over one steel rail in between the initial and the destination rooms. A metal/plastic peg (approx. 20cm long, diameter: 1.5cm) served to prevent the door to be open without being removed first. To avoid that chimpanzees took it out, it had a screw that would move along a horizontal incision and stopped, thus allowing the door to slide (to open) while keeping the peg attached to the door structure.. Figure 27 shows the sliding mesh door built up over the steel structure that separates both rooms. The door is half open and the chimpanzee is sitting on the rail. The peg has been manipulated to let the door open but it has also been bent.

Figure 27. Door structure. Daza observes the metal peg she has bent.

In this phase we also used food to encourage the chimpanzees opening the door and moving to the destination room. We used grapes, slices of bananas, onions, pellets and yoghurt, according to every subject's preference. We also had one camera mounted on a tripod and coding sheets.

Exposure

During the exposure the subjects see how some experimenters interacting with them (Study 3) or with a conspecific (Study 4). For both studies we needed a sliding table (approx. 70cm x 35cm) that would stand over a metal structure and a plastic stool to sit

down. For Study 4 we also used a juice dispenser that we held with tape on the crystal glass of the experimental room and whose plastic mouth was inserted in a hole on the safety panel. The juice dispenser was a hospital drip that we filled with diluted grape juice. We recorded the sessions with two cameras (plus tripods).

The food we used were half grapes that the experimenters hide in/take from/get into their front pockets.

Punishment

For this phase we needed to occlude part of the experimental room. The elements we used to occlude were: black plastic curtain hanging from the ceiling (approx. 1.50cm x 70cm) (see Figure 28, upper left); pieces of PVC that we fixed to the mesh with screws (see Figure 29) and black tape to cover whatever remaining visual access between the mesh the subjects could have. To record the test we needed five cameras (for disposition of the cameras see Figure 34 in Procedure) with their corresponding tripods, one flexible structure with a handle that could be attached to the safety panels, five cables and a splitter (see Figure 28 upper right) to merge the output of each camera. The object that the punisher used to hit was a green water noodle (approx. 60cm) (see Figure 28 right). During the whole phase we used a metronome to help the punisher adjust the hits to the rhythm so that all the punishment phases lasted the same.

Preference

During the preference phase the experimenters entered in the room with grapes. In Study 3 each used a bucket full of food, whereas in Study 4 the experimenters used a transparent plastic box with 20 grapes inside. We used two plastic stools to sit down during this phase. An experimenter outside the room had a stopwatch to count 2 minutes.

Figure 28. Material for Studies 3 and 4 (chimps). Black curtain, tape, plastic stool, tripods (upper left); five cameras, screws, plastic tuppers, flexible camera-handle, coding sheets, chronometer (upper right); green water noodle (right).

Figure 29. Occluders used to prevent chimpanzees from watching the punishment unless they opened the sliding door.

Studies 3 and 4 (chimpanzees): procedure

Training

In the training the chimpanzees learnt how to open a heavy metal door by manipulating a metal/plastic peg. This manipulation required certain physical effort. By opening the door, the chimpanzee could get access to a different room from which he/she would be able to continue watching the actions of the experimenters. Therefore, the manipulation of the door was the equivalent to placing the coins in different boxes that we used for children: costly action to show interest for witnessing a punishment.

The training on how to open the door comprised of 5 different consecutive stages of progressive difficulty. The mesh sliding door was positioned in between two rooms (start and destination rooms) of the chimpanzees. The objective of every stage was to open the door. To encourage subjects to do so, slices of bananas (and sometimes pellets, grapes or slices of onions, depending on individual preferences) were placed in the destination room. Subjects could witness the baiting process before the trial started and the food was visible to them all the time as the wall in between the rooms was made out of mesh. In case subjects did not attempt to open the door within 5 min., the respective trial was re-started. Subjects from chimpanzee group A (see Table 23) had already received a training to open the sliding door in a previous experiment (Bullinger, Burkart, Melis, & Tomasello, 2013) and therefore only received refreshment trials (stage III) two years after the initial training.

Stage I. The sliding door was left 1/3 open. Subjects had to slightly push the door to the side in order to get into the destination room and retrieve the food. The criterion for advancing a subject to the next training stage consisted of passing through the door without showing signs of fear. It took subjects from group B an average of 2 trials (range, 1–4 trials) to reach the criterion for this stage. One subject (Alexandra) dropped out as she never reached the criterion (showed signs of fear in crossing the door) even after 50 trials.

Stage II. The door was completely closed but not locked. That is, subjects had to pull the door to the side but did not have to unlock it. As in stage I, the criterion for advancing a subject to the next stage consisted of opening the door and passing through it without showing signs of fear. It took subjects from group B an average of 1.6 trials (range, 1–3 trials) to reach the criterion for this stage. One subject (Annett) dropped out

as never reached the criterion (showed signs of fear in crossing the door) even after 25 trials.

Stage III. The door was not only completely closed but also locked with the help of a peg. Initially, we used a removable plastic peg that we replaced by a fixed metal peg. The plastic peg blocked the opening of the door, but once it was taken out the door could be pulled to the side just like in the previous stage (stage II). In case subjects did not realize that the peg had to be removed in order to open the door, the experimenter encouraged the exchange of the peg with food items. This procedure was facilitated for some subjects by placing similar pegs to the one blocking the sliding door on the floor of the room. As soon as the subjects learnt how to unlock the door, the plastic peg was replaced by a metal peg that could not be detached from the door. The criterion for advancing a subject to the next phase consisted of opening the sliding door and moving into the destination room within 10 sec. in four consecutive trials. It took subjects from group B an average of 41 trials (range, 9–103 trials) to reach the criterion for this stage. Subjects from group A took an average of 11 trials (range, 6-24 trials) to learn how to open the door. In Figure 30 we show images of how this procedure worked.

Stage IV. No food was placed in the destination room and the door was closed and locked with the metal peg. Two of these trials were included in order to eliminate the association between opening the door and receiving food.

Figure 30. Opening door process by chimpanzees: initial state of the door (upper left); unlocking the door by pulling from the peg (upper right); pulling the door to open it (bottom left); moving into destination room (bottom right).

Exposure phase

This phase varied depending on the study (direct harm-study 3; indirect harm-study 4). In Study 3 the chimpanzees could watch punishment after having directly experienced pro- and antisocial behaviors. In Study 4 the chimpanzees were bystanders of how a conspecific experienced pro- and antisocial behaviors and then they could decide whether watching the experimenters being punished. Therefore to run this phase in Study 4 we needed to keep the attention of the subject constant to the interaction between experimenter and conspecific.

In Study 3, one of the experimenters (the starting order between pro- and antisocial was counterbalanced) came into the testing room alone and sat in front of a sliding table attached to the subject's room. The experimenter placed half a grape on top of the sliding table and moved it towards the subject. As soon as the subject tried to reach for the food, the prosocial experimenter allowed it to take it, while the antisocial experimenter brought back the sliding table thus preventing the subject from taking the food. Within one session each subject received three blocks of 12 trials each (6 trials with each of the two experimenters). The experimenters counterbalanced their role across subjects. This phase took place one day before the Punishment phase. In addition, two refresh trials were administered the day of the punishment phase, right before the actual test.

In Study 4 each of the experimenters proceeded in the same way as in Study 3 but towards a stooge and not towards the subject. The subject was in an adjacent room and witnessed the interaction between the agent and the stooge. To ensure that the subject watched the whole interaction a juice dispenser was added to the booth area of the room. The spout of the dispenser was located in such an angle that while drinking subjects directly faced the experimenter and the stooge. Each of the experimenters made sure that every trial was watched by the subject. That is, they looked sideways to the subject and in case he/she stopped drinking, they waited to restart the interaction with the stooge until the subject was again gazing at them (see Figure 31).

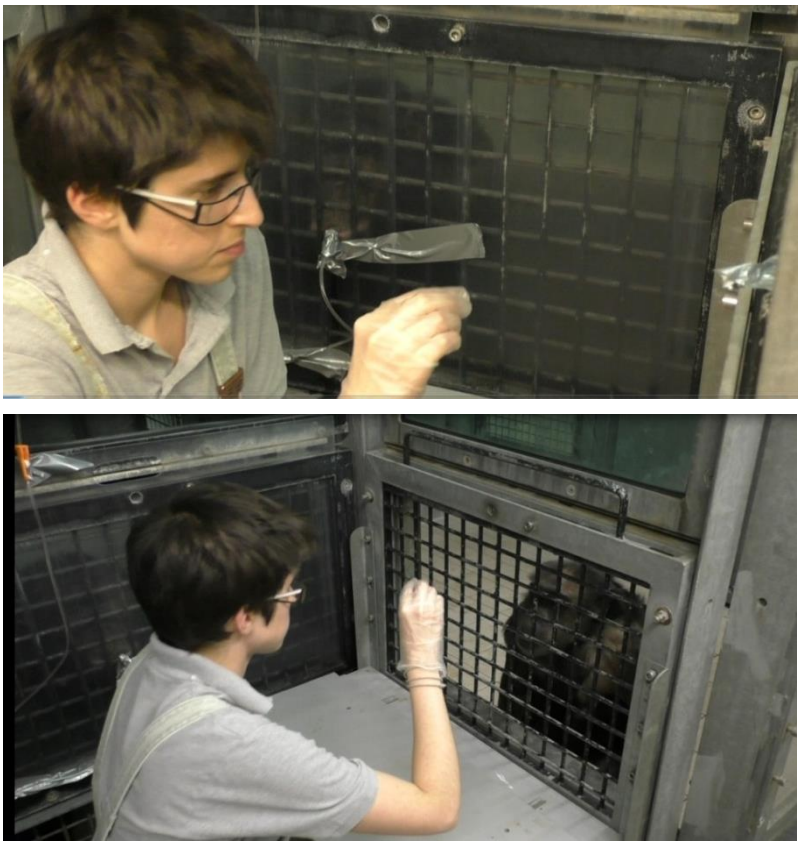


Figure 31. Exposure phase in Study 4. The subject witnesses the interaction experimenter-conspecific while drinking from a juice dispenser (subject view-upper; stooge view-bottom).

Preference phase

As we did with the children, in this phase we wanted to be sure that the subjects had understood the role of the experimenters. In order to know that, we let them choose who experimenter to approach and beg food from, although the experimenters never responded to their gestures (i.e. did not give food). We should expect that chimpanzees approached and begged more from prosocial experimenters. In Study 3 we conducted this phase right before the punishment phase. However in Study 4 we run the punishment first and the preference phase afterwards. This was done to keep the role of the agents unaffected. Since in Study 4 the interaction with the experimenter was indirect (the experimenter had been prosocial or antisocial towards a third-party), if subjects begged from the experimenters and they did not respond to them that could turn into a direct harm polluting the indirect nature of the Study.

In Study 3, both experimenters entered the testing room simultaneously, each carrying a bucket with equal amount of grapes. Both experimenters sat in front of the subject's room with their backs oriented towards each other (see Figure 32, left). The agents remained in their position for 5 min. from the moment the subject came into the room.

As in Study 3 chimpanzees did not show a preference for the prosocial experimenter and this could have been the result of the sub-optimal spatial arrangement of the experimenters that might have caused some confusion because the chimpanzees could not see both experimenters simultaneously, in Study 4 we followed a previous set-up (Russell, Call, & Dunbar, 2008) that had proved to yield clear choices when apes were asked to decide between social. Therefore in Study 4 the preference phase was very similar to the Study 3, except for the fact that both agents sat side by side, each of them in front of a different mesh panel (see Figure 32, right).

Figure 32. Preference phases. Here we displayed how experimenters sat down in the preference phase for Study 3 (left) and Study 4 (right).

For this phase the following vocalizations and other behavioral signs of display, distress and begging gestures were coded from the usb video cards, respectively: 1) spat, bang, throw feces at experimenter, knock/shake window panel, hoots; 2) scratch body/nose, whimper, scream; and 3) loose under-lip, boob-head-body, present bottom at experimenter, poke through the holes in the window panel and/or grab bucket from experimenter. The duration of the calls was analyzed with the sound analysis software Avisoft ® and Praat ®.

Punishment phase

We needed one hour and a half to set-up this phase and half an hour to unset-up. This is because immediately before running the punishment trials we conducted two refreshment exposure trials, for which we needed to set and unset-up the sliding table, the juice dispenser (in Study 4) and the cameras. The specific setting up for the punishment phase entailed to occlude the whole mesh of the experimental room; placing the metal peg in the sliding door; switching off the metronome plus relocating the five cameras (see Figure 34); connecting them with cables to the splitter and confirm that they all were working before starting.

In this phase we wanted to see whether chimpanzees were willing to incur a cost to continue watching punishments, and if they did it more frequently with pro- or antisocial experimenters. As we did with the children, a third character was introduced, the punisher, who remained the same across all conditions and subjects (i.e. Katrin Schumann). Since we had no curtain to lower of, we run two different conditions: visible and invisible. In the visible condition, the subjects had full visual access to the punishment. If the subjects wanted to escape from the punishment happening in front of them, they had to operate the door to move to another part of the room where this would then be out of sight. In the invisible condition we occluded part of the experimental room with black material and PVC pieces, preventing the subjects to see what happened behind them (see Figure 29). In case they wanted to continue watching, they should move to another room. To do so, they should incur the cost of opening the door as they had learnt during the training. We ran four punishment trials per subject (visible-prosocial; visible-antisocial; invisible-prosocial; invisible-antisocial) in two consecutive days. Each day we ran two trials counterbalancing the condition (eg. visible-prosocial and invisible-antisocial). To ensure that the duration of the punishment was equal across conditions, we had a metronome on so that the punisher made the hits coincide with the metronome's rhythm.

Both conditions started with the particular experimenter entering in the room and sitting down on a plastic stool during 5 seconds (we referred to this period as "baseline" in the Results section). After that time, the punisher entered in the room with a water noodle in her hand (the equivalent to the stick in the child's experiment). The punisher approached the experimenter from behind with a human facial expression of rage (we referred to this period as "pre-hit" in the Results section) and started beating her up (we referred to this period as "hit" in the Results section) with the water noodle for 4 sec. (i.e., 4 hits with the stick, rate: 1 Hz). While being beaten up the agent cried out in pain. After the initial punishment period (i.e., 4 seconds) the experimenter either: 1) remained in her initial position (area A, see Figure 33 and 35 visible condition) for the whole time of the punishment (10 more seconds), so-called visible condition, or 2) left her initial position and went into another area of the room invisible to the chimp (area B, see Figure 33 and 35 invisible condition) where the punishment continued for 10 more seconds, so-called invisible condition. As in the invisible condition the experimenters needed 5 sec. to move into another part of the room, to ensure that the duration of the

punishment was equal across conditions, in the visible condition, after the fourth hit with the stick, the punisher brought it further back, to achieve an even stronger hit, for 5 sec.

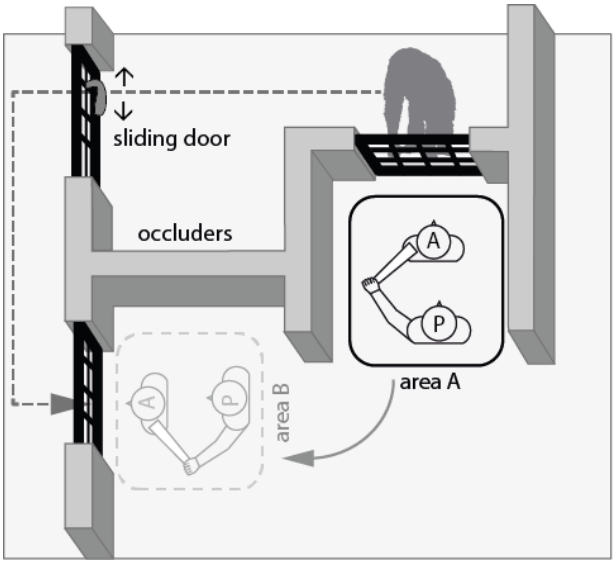


Figure 33. Procedure for Studies 3 and 4 (chimpanzees). Here we displayed where the experimenters acted for each condition: visible (area A) and invisible (first area A, then move to area B).

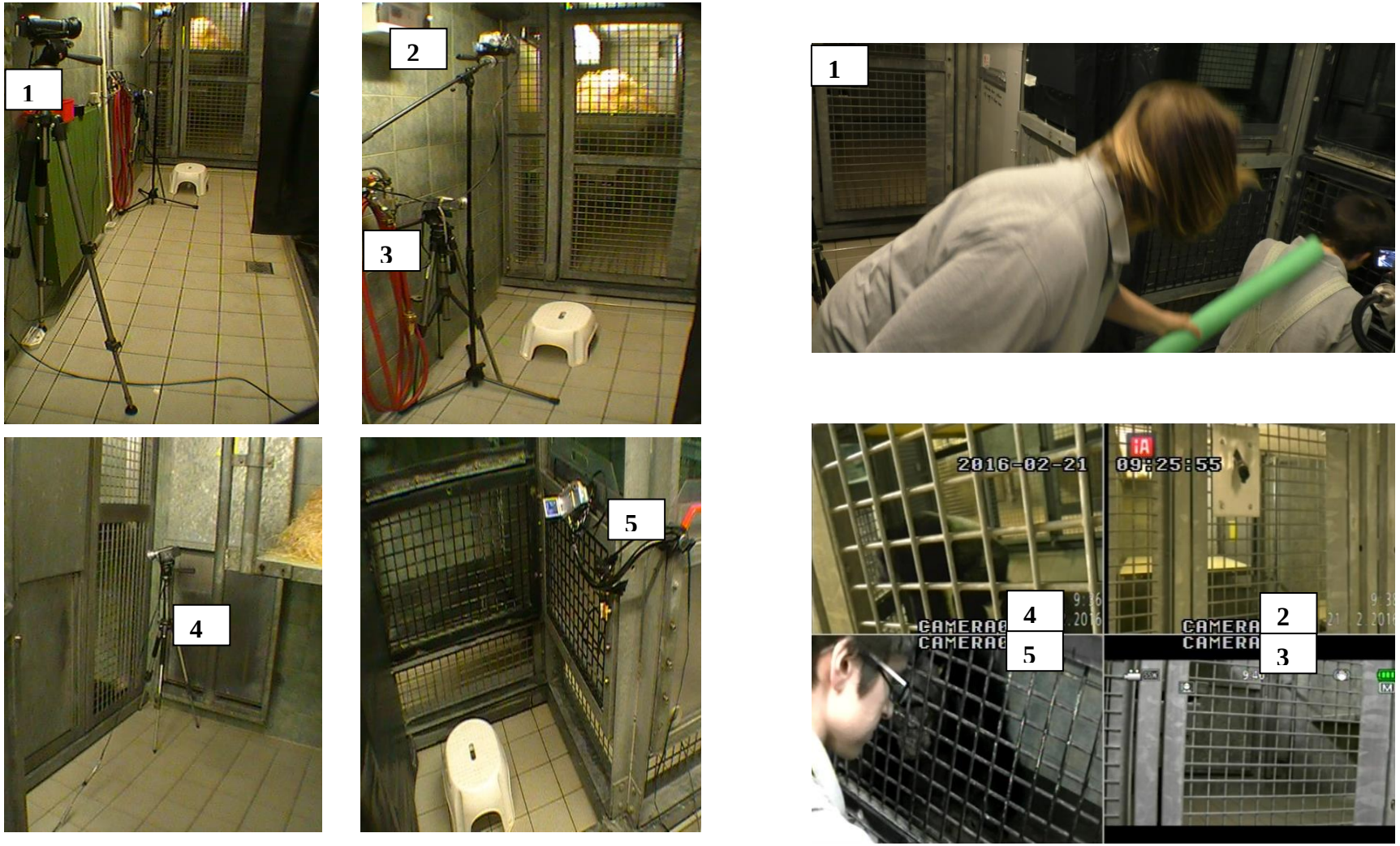


Figure 34. Disposition of the 5 cameras (left) and outcome videos after testing (right).

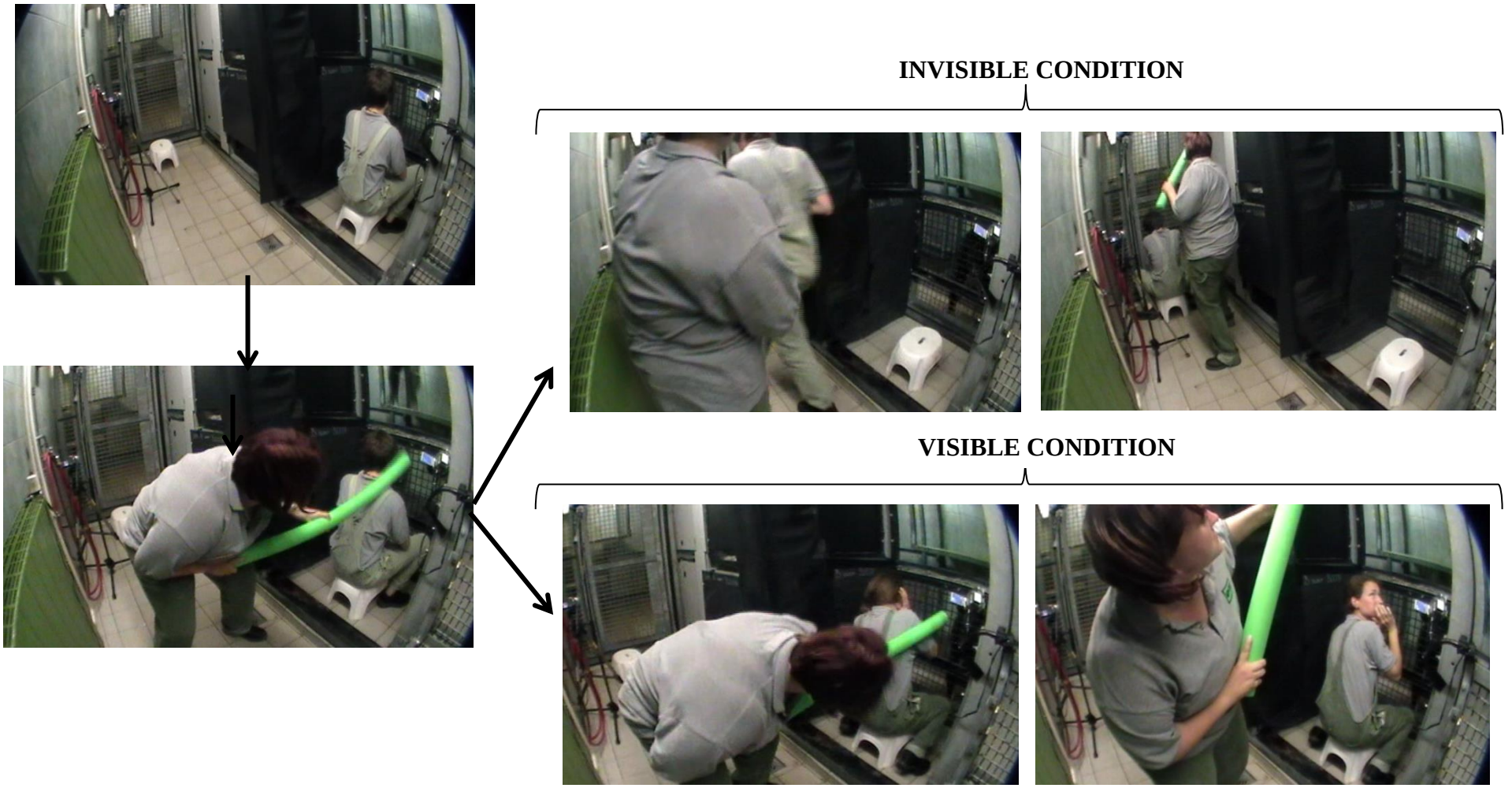


Figure 35. Procedure for visible/invisible conditions with chimpanzees. From left to right: first, the experimenter enters. Then, the punisher enters and starts hitting her. After that, the procedure changes between visible and invisible conditions. In invisible condition (upper), both experimenters moved to another part of the room. In the visible condition (bottom), both experimenters stay.

Analysis

We conducted two studies with chimpanzees. Study 3 with chimpanzees was comparable to Study 3 with children (i.e. used same compatible procedure and design) and it explored whether chimpanzees were willing to incur costs to watch punishment or previous pro- or antisocial agents. Therefore, we analyzed whether there were differences between the times they opened the door (i.e. incurred cost) to watch the prosocial and the antisocial agent (see "Schadenfreude" below). Besides, we had included a different condition from that of children (i.e. visible condition) to see whether chimpanzees empathized with the prosocial agent and wanted to escape while watching her punishment more than when watching the antisocial. Therefore, we looked for differences between the times they opened the door to escape from watching the punishment of pro- and antisocial agents (see "Empathic distress" below). It is crucial to note here that our dependent behavioral variable of opening the heavy sliding door for the chimpanzees was the same throughout all conditions. However, we interpret it differently depending on the condition (i.e., to continue witnessing the punishment when it is invisible or to escape into another room when it is visible).

Importantly, incurring a cost to watch an antisocial agent being punished might merely reflect that this is seen as more socially informative or coherently into the flow of the preceding events. That is why we also ran Study 4. There chimpanzees did not directly experience but witnessed how the prosocial and the antisocial agents interacted with a conspecific (stooge). Our analysis was the same for the invisible and visible conditions (i.e. looking for differences in the number of times that the door was operated). All our comparisons were two-tailed except for the prosocial condition.

Finally, in each study we analyzed their arousal during the initial punishment to see whether it changed depending on the agent (prosocial or antisocial). We collapsed the vocalizations of both visible and invisible conditions because the arousal for watching the prosocial or the antisocial should be the same regardless of the condition. To do so, we divided the punishment event into three periods: an initial baseline where just the agent was present; a pre-hit period where the punisher appeared but had not started to punish the agent, and a first-hit period during which the punishment actually took place. We looked at these periods separately for each of the two agents and categorized the vocalizations according to their acoustic and temporal properties (Nishida, Zamma,

Matsusaka, Inaba, & McGrew, 2010) and grouped according to the call categories suggested by Goodall (Goodall, 1986): screams, whimpers, and worried hoos considered as distress vocalizations and (waa) barks and (pant) hoots considered as display vocalizations. Distress and display vocalizations were lumped together and the combined results used for statistical analysis. The duration of the calls was analyzed with the sound analysis software Avisoft ® and Praat ®.

Studies 3 and 4 (chimpanzees): results

Results

Interobserver reliability

To assess inter-observer reliability[43], a second observer coded a random sample of 20% of the trials. Inter-observer reliability was high for duration of the vocalizations (Study 3, distress calls: r = 1.000, P < 0.001; display calls: r = 0.900, P < 0.001; Study 4, distress calls: r = 1.000, P < 0.001; display calls: r = 1.000, P < 0.001) and for opening the sliding door (Study 3: Pearson correlation r = 1.000, P < 0.001; Study 4: r = 1.000, p < 0.001).

General comprehension

We looked at the preference phases in each study to see whether chimpanzees understood the different roles of the agents (prosocial and antisocial). We analyzed the frequency of begs corrected for the amount of time spent in front of the correspondent agent (Study 3: r = 0.999, P= 0.028; Study 4: r = 0.997, p = 0.048).

Study 3 (direct experience)

Preference phase

Chimpanzees showed no preference for requesting food from the prosocial over the antisocial agent in Study 3 (Wilcoxon signed rank test: T+ = 89, N = 17, P = 0.579). This could have been the result of the close physical proximity of both agents, which might not have allowed for a clear dissociation of the subject's behavior. Plus, the scenario was very confusing (two experimenters sitting back to back forcing the chimpanzee to move from one room to another depending on the human he wanted to beg from). This was the reason why in Study 4 we changed the physical disposition of the experimenters in this phase. With the simple dispositional change the preference was very clear even when the moral reputation of the agents had been learnt indirectly (see below). Therefore, we could assume that the low score for the preference phase in

[43] The coding of the vocalizations was a very tiring task due to the quality of the sound in the videos. I want to specially thank Yseult Hejja-Brichard and Katrin Schumann for helping with the analysis of the chimpanzee vocalizations and Markus Neuschulz and Anja Hutschenreiter for reliability with the chimpanzee data. We also received help to identifying vocalizations and distinguishing their different types from Thibaud Gruber, Catherine Crockford, and Ammie Kalan (who also showed us how to use the software Avisoft ® and Praat ®).

Study 3 was due to the way the experimenters sat down and not due to lack of comprehension, and continued the analysis.

Schadenfreude

Chimpanzees differentially operated the heavy sliding door depending on whether punishment was visible or not and whether the agent had been previously prosocial towards them or not (Cochran's Q = 8.59, df = 3, P = 0.043, N = 16). We conducted pair-wise follow-up comparisons between the two invisible conditions to test our hypothesis of an increased motivation to witness the punishment of an agent who had been previously antisocial towards the subject. Subjects were significantly more likely to incur the physical costs to open the heavy metal door in the antisocial invisible condition (50% of the subjects) compared to the prosocial invisible condition (18.75% of the subjects) (Sign test: P = 0.032, N = 16, one-tailed; Figure 36, left graph-left bars).

Empathic distress

We conducted another pair-wise follow-up comparison between the two visible conditions to test for the behavioral effects of empathic distress (i.e. increased opening of the door to move to another room when the punishment of the prosocial agent is visible to the subject). Here we found no significant difference in the number of subjects who opened the door during the prosocial visible condition compared to the antisocial visible condition (Sign test: P = 0.313, N = 16, one-tailed; Figure 36, left graph-right bars).

Vocalizations

There was a significant difference between the three periods in the duration of the vocalizations in the presence of the prosocial agent (Friedman exact test: F = 9.82, P = 0.004, N = 16; Figure 36, right graph) but we found no such difference in the presence of the antisocial agent (F = 4.67, P = 0.107, N= 16; Figure 36, right graph). Comparing the vocalizations in response to the presence and punishment of the prosocial and the

antisocial agents, showed that chimpanzees produced longer vocalizations in the baseline period when facing the antisocial agent compared to the prosocial one (Wilcoxon exact test: T+ = 21, P = 0.031, N total = 16; corrected for the duration of each period in the Punishment phase, i.e., baseline, pre-hit, hit periods) and longer

vocalizations when the prosocial agent was being punished compared to when the antisocial agent was being punished in the hit period (Wilcoxon exact test: T+ = 21, P = 0.031, N total = 16; Figure 36, right graph).

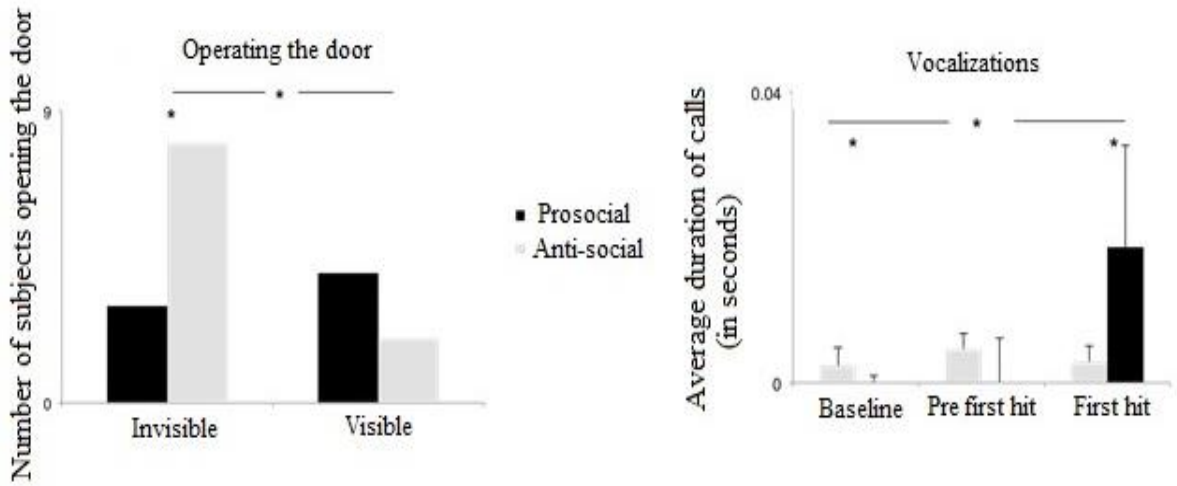


Figure 36. Schadenfreude results in chimpanzees, i.e. number of subjects who opened the door in the invisible condition (left graph, bars on the left). Empathic distress results, i.e. number of subjects who opened the door in the visible condition (left graph, bars on the right). Vocalization results, i.e. average duration of calls during visible and invisible conditions by valence of the agent (right graph).

As a summary, more chimpanzees opened the heavy sliding door to continue watching the punishment in the invisible antisocial compared to the invisible prosocial condition. Note, that not all the chimpanzees opened the door. When watching the punishment of the prosocial agent there were no signs of empathic distress in the form of opening the door to escape watching the scene but chimpanzees did express greater distress vocalizations than while watching antisocial agents being punished.

Study 4 (indirect experience)

Preference

We used Russell et al.'s paradigm to test for a potential preference between the prosocial and the antisocial agent (Russell et al., 2008). We found that chimpanzees begged significantly more often from the prosocial than the antisocial agent (frequency of begs corrected for the amount of time spent in front of the correspondent agent, Wilcoxon signed rank test: T+ = 82, N total = 14, P = 0.008).

Schadenfreude and Empathic distress

Unlike Study 3, we found no evidence that chimpanzees differentially opened the heavy sliding door in the four conditions (Cochran's Q = 3, df = 3, P = 0.484, N = 14).

Vocalizations

There was no significant difference between the three periods in the duration of the vocalizations in the presence of the prosocial and antisocial agents (prosocial, Friedman exact test: F = 0.125, P = 1.00, N = 14; antisocial, F = 3.26, P = 0.218, N = 14).

As a summary, when chimpanzees indirectly experienced the niceness or nastiness of a third-party they were not willing to incur costs to watch the nasty being punished (i.e. Schadenfreude) not to escape from watching the prosocial being punished (i.e. empathic distress). Moreover, their vocalizations did not vary from one scene or the other.

We summarize all the results from Studies 3 and 4 in Table 24 on the next page.

Table 24.
Summary of the results of Studies 3 and 4.

Concept	Key points
Preference	Children had a preference for the prosocial agent.
	Study 3: Chimpanzees did not show clear preference for the prosocial (but methods shortcoming)
	Study 4: Chimpanzees had a preference for the prosocial agent.
Indirect revenge (if second-party)	6-year-olds and chimpanzees incurred significantly more cost to watch the antisocial punishment
	6-year-olds showed a mixture of positive and negative emotions during the antisocial punishment
	Chimpanzees produced longer vocalizations during the prosocial punishment
Indirect revenge (if third-party)	Chimpanzees did not show consistent preference for watching punishment, regardless of the agent
	Chimpanzees did not vocalize differently in any punishment, regardless of the agent

Studies 3 and 4 (children and chimpanzees): discussion

Discussion

Our findings demonstrate that chimpanzees and 6-year-old but not 4- and 5-year-old children appear to possess a motivation to watch the punishment of others who they had previously experienced as antisocial towards themselves as compared to prosocial agents. Thus, chimpanzees endured greater physical efforts and 6-year-old children spent more valuable coins to continue watching the punishment of an agent who had previously withheld something valuable from them (i.e. food for the chimpanzees and favorite toys for the children) as compared to someone who had been prosocial and shared the valuable items. In contrast, chimpanzees spent the same effort to continue watching the punishment of an agent regardless of the agents' social inclination when these were not directly involved but just passively observed such interaction between two other parties.

We also observed concomitant indicators of affective responses in the children. Thus, 6-year-old children showed a greater mixture of positive and negative emotions in response to watching the punishment of the antisocial agent. The combination of these emotions lies at the core of an experience induced by watching something unpleasant occurring to someone who deserved it while simultaneously rejoicing in it, a feeling also known as Schadenfreude (Smith et al., 1996). There is some prior evidence that around 6 years of age, children are capable of experiencing such potentially conflicting emotions (Steinbeis & Singer, 2013). In the present context, however, it is unclear whether the emotional experience is a cause or a consequence of continuing to watch the punishment. In addition to such signs of Schadenfreude, we found signs of empathic distress in children of all ages. However, this was not differentially modulated by whether the agent had been previously prosocial or antisocial towards them. Even though children as young as 3 years old have been shown to differentiate their empathic helping between previously prosocial and antisocial others (Eisenberg et al., 1994; Vaish, Carpenter, & Tomasello, 2010), such a difference was not observed in the present study.

Interestingly, chimpanzees produced longer vocalizations indicative of emotional arousal during the punishment of the prosocial agent that had directly interacted with them but no differential vocalizations occurred when they witnessed the agent being punished following the indirectly experienced pro- and antisocial behavior (regardless

of her social orientation). Even though in chimpanzees it is difficult to clearly label the valence of such vocalizations as they can reflect conflicting emotions (Goodall, 1986), the specificity of their occurrence (longer vocalizations during the hitting of the prosocial agent compared to the antisocial agent) suggests that they might reflect something akin to empathic distress. However, chimpanzees did not signal distress by attempting to escape witnessing the punishment of the prosocial agent.

Studies have shown that chimpanzees engage in punishment of conspecifics that had previously stolen their food by causing the thief's food to disappear (Jensen et al., 2007b; Riedl et al., 2012). Our direct Schadenfreude experiment with chimpanzees demonstrates that also in the absence of food, chimpanzees are motivated to watch antisocial agents being punished after directly experiencing the antisocial behavior themselves. One could argue that the chimpanzees' reaction could be driven by emotional engagement. However, chimpanzees were more aroused when they watched punishment of the prosocial agent. Following indirectly experienced pro- and antisocial behavior chimpanzees were equally motivated to watch punishment of the pro- and the antisocial agents. This is consistent with findings showing that chimpanzees do not punish those who stole food from third parties (Riedl et al., 2012). The results from our indirect Schadenfreude experiment, in which chimpanzees merely observed the prosocial and antisocial interaction prior to the agents' punishment, help us to interpret the results from the direct Schadenfreude in which chimpanzees engaged directly with the prosocial and antisocial agents. In both studies all basic elements were kept constant except for the degree of the chimpanzee's involvement. Thus, alternative explanations such as increased social informational value or a greater coherence in the unfolding of the scene to account for the increased motivation of observing punishment of antisocial agents following a direct interaction can be ruled out. Instead, the most likely interpretation based on these findings is that chimpanzees have an increased motivation to observe such punishment because it follows a desirable action towards someone who behaved antisocially towards themselves. The literature abounds with examples of animals willing to incur energy costs for something they find rewarding (eg. Beran & Evans, 2009). It is therefore tempting to argue that watching antisocial others getting harmed is rewarding and pleasurable also to chimpanzees. However, in the absence of direct evidence for the presence of such positive emotions, we remain cautious with such an account.

Although children of all ages incurred costs to watch some of the punishment, only 6-year-olds (but not 4- and 5-year-olds) were willing to differentially incur greater costs to continue watching the punishment of the previously antisocial agent. This was the case despite a decided preference for the prosocial agent also in younger children. Further, only 6-year-olds showed a greater number of combined positive and negative emotions when watching the initial punishment of the antisocial agent compared to the prosocial one. These data suggest that in children, pleasure at seeing deserved punishment may be linked to the increased costs incurred to continue watching it. Recent studies have shown that differential punishment of selfish behaviors of in-group and out-group members already occurs from 6 years onwards (Jordan et al., 2014). This suggests that 6 years of age may be a critical developmental time point at which children are willing to actually sacrifice their resources to see fairness enacted (McAuliffe, Jordan, & Warneken, 2015). Surprisingly, four-year old children showed a greater willingness than the 5- and 6-year olds to give up stickers in order to continue watching punishment of the puppets regardless of how the agents had previously behaved. Stickers are valuable items for children of this age (eg. Bueno-Guerra et al., 2016; Engelmann et al., 2013; Smith et al., 2013) and we believe that this indiscriminate use reflects their overall enjoyment of the puppet performance. There are some limitations to the present set of studies. One is the fact that interactions were observed between individuals that were not of the same species as the subject. However, this concern is reduced given that both chimpanzees and 6-year-olds responded differentially to the two agents. While such cross-species set-ups are common in the study of social behavior of both human and non-human primates (eg. Call, Hare, Carpenter, & Tomasello, 2004; Whiten, Custance, Gomez, Teixidor, & Bard, 1996) future work will have to assess how far these findings extend onto interactions with one's own species. Further, the different dependent variables for the chimpanzees and the children (i.e. physical energy vs. valuable coins) make direct interspecific comparisons difficult. While using different dependent variables has the advantage of optimizing procedures for each species thus avoiding potential biases favoring one of the species, future work may seek to expand the findings using the same dependent variables for greater comparability of the effects. Finally, we were unable to counterbalance the administration of the direct and indirect exposures to the pro- and antisocial in chimpanzees. Our results, however, were consistent with the existing literature on the occurrence of punishment following

directly and indirectly experienced transgressions in chimpanzees, which ameliorates to some extent the concerns derived from our current design.

We studied the evolutionary and ontogenetic origins of an increased motivation to watch the punishment of antisocial others and their associated emotional states. Chimpanzees and 6-year-old children showed greater motivation by incurring costs to continue watching the punishment of an antisocial over a prosocial agent. Furthermore, children displayed differential responses of mixed positive and negative emotions when they witnessed punishment of antisocial agents, which suggest that they might take some form of pleasure from this. Although such a mechanism is still uncertain in chimpanzees, vocalizations of emotional arousal produced when they witnessed the suffering of a prosocial agent. Their absence when witnessing the suffering of an antisocial agent might indicate that affective responses such as pleasure may constitute an important motivational contributor to the exaction of revenge with early evolutionary origins. Crucially, chimpanzees did not vocalize differentially for the two agents when seeing the two agents punished following indirectly experienced pro- and antisocial behavior. Additionally, they did not engage in differential costs to witness the punishment of the antisocial agent as compared to the prosocial agent. These findings provide some evidence for the evolutionary origins of an increased motivation to watch punishment of antisocial behavior with - at least in children- possible links to feelings of pleasure underlying such a motivation. Such a motivation appears to develop at a protracted rate, similar to higher level cognitive skills (Hanus, Mendes, Tennie, & Call, 2011) and might emerge at an age at which children begin to care so much for abstract entitlements such as justice that they are willing to pay for it.

General discussion

General discussion

The purpose of this dissertation has been to contribute to the evolutionary and ontogenetic study of punishment by exploring the existence of functional punishment in chimpanzees (Study 1); functional spite (i.e. revenge) in children (Study 2) and feelings associated to indirect revenge in chimpanzees and children both in second-party (Study 3) and third-party scenarios (Study 4).

With regard to Study 1, our main finding is that chimpanzees, contrary to humans, do not consistently use functional punishment in iterated tasks to foster cooperation. In fact, rejections seem to cause the opposite behavior, a decrease in generosity. We have proved this by showing that chimpanzee responders never rejected non-zero outcomes even when the offered distribution caused inequality between the players. Additionally, we have shown that chimpanzee proposers reduced their prosocial offers across sessions in an MUG, as opposed to in a DG, probably due to the emergence of frustration after zero-offers rejections. The emergence of frustration prevents future cooperation to arise and reinforce the persistence of non prosocial offers. This might be the cause why cooperation trough punishment does not work in chimpanzees: while humans interpret punishment as an invitation to change their behavior to be adapted to the partner's needs, chimpanzees might interpret it as a food loss that only produces anger and consequently lowers their disposition to offer generously.

With regard to Study 2, our main finding is that children show functional spite from very young but this punishment is mainly based in the disparity of outcomes, without the interplay of their partner's intentions. Besides, child proposers do not seem to be able to anticipate functional spite and tend to offer selfishly, with the exception of older girls. We have proved this by showing that there were 8/2 rejections in a one-shot MUG both at 6- and 10-years of age without significant differences between conditions. Despite in some conditions 8/2 was the best alternative to offer, children rejected it indiscriminately, especially 10-year-olds. This age group showed high levels of disadvantageous IA. Additionally, proposers did not consistently offer what was desirable for same age peers, showing low competence to form a priori cooperative interactions. The exception was 10-year-olds girls, who were able to offer according to their partners' expectances.

With regard to Study 3, our main finding is that both chimpanzees and 6-year-olds are willing to incur costs to watching previously antisocial agents being punished. Furthermore, both children and chimpanzees showed emotional signs of Schadenfreude and empathic distress, respectively. We have proved this by showing that both species incurred cost (in the form of coins or physical effort while operating the door) significantly more in the invisible antisocial condition than in the invisible prosocial condition. Additionally, at the emotional level, 6-year-olds showed mixture of positive and negative signs (co-occurring frowns and smiles) during the antisocial punishment and chimpanzees produced longer vocalizations during the prosocial punishment.

Finally, with regard to Study 4, our main finding is that chimpanzees are not willing to incur cost for events that did not directly happen to them. We have proved this by showing that chimpanzees did not incur cost nor produce any emotional signs significantly more in one condition over other.

Next, we would like to discuss how all these results shed light upon the evolutionary origins of punishment. To do so, first we will focus on the differences and similarities between humans and chimpanzees when bargaining (studies 1 and 2), with special attention to the ecological validity of the methodological paradigm employed. After that, we will do the same for studies 3 and 4. Finally, we will provide a last paragraph trying to summarize the most meaningful key points of the present dissertation.

We started this dissertation distinguishing between different types of punishment (i.e. functional punishment, functional spite and indirect revenge), discussing their functions and wondering whether some of the types of punishment that we already know that exist in humans, may also exist in other primates. Darwin seemed to be very confident about the later point. In his book *The descent of man*, he stated that revenge (i.e. functional spite) was common through all primates and even extensible to some other taxa:

> "The man and the higher animals, especially the Primates, have some few instincts in common. All have the same senses, intuitions and sensations-similar passions, affections and emotions, even the more complex ones (…) they practise deceit and are revengeful"(Darwin, 1871, p. 54)

However, this assertion had not been empirically tested until very recent. One of the first (if not the pioneer) experimental studies about punishment in apes was conducted by Jensen and colleagues in 2007 when they showed that chimpanzees punished the theft of food (Jensen et al., 2007b). This was the first empirical evidence of non-human primates executing some sort of punishment in an experimental setting: functional spite (taking revenge). However, the same year, the same chimpanzees did not reject unequal offers in the ultimatum game (Jensen et al., 2007a), nor even when there was also theft involved (Riedl et al., 2012). This absence of rejection was also present in different chimpanzee subjects playing the same game with different methods (Proctor et al., 2013a) and we have also shown that in our Study 1, even when we were looking for functional punishment. What is the explanation for this? Is it that punishment is a very human thing or is it, maybe, that the different methodologies employed preclude to the accurate study of punishment in chimpanzees?

After all the experiments conducted so far, we believe that the key point is the type of paradigm used to test punishment. Note that in Jensen and colleagues' first experiment chimpanzee subjects were peacefully enjoying some food when, all of a sudden, a conspecific took it away from them. The partner became an offender because he made some current possession disappear. By contrast, in the ultimatum game experiments, the subjects did not enjoy anything a priori but they receive something from their partners. Why should then the partner become an offender? Usually, chimpanzees in the wild do not offer food to each other (Gilby, 2006), so, independently to the partner's outcome, any offer is a present and thus no offense is perceived. Besides, the cost to be incurred in the first experiment in order to protest is some physical effort, without any material lost. By contrast, in the ultimatum games the cost to be incurred is food. Chimpanzees show such a marked preference for food that it is the main reward used for the majority of primate experiments. In fact, chimpanzees are able to incur surprising tiring efforts just for a little piece of food like a peanut (remember the floating peanut task (Hanus et al., 2011)). If they are willing to take minutes to go for one place to another repeatedly just to obtain some food, how are we expecting that they could ever reject windfall available food? Therefore, it might happen that the problem when trying to study punishment in chimpanzees (regardless whether it is functional punishment or spite) is the use of some paradigm that works for humans but that it is not ecologically valid for chimpanzees.

The concept of umwelt coined by Uexküll (von Uexküll, 1920) refers to how animals see the world, with "see" being a compendium of their motivations, perceptual and frequent desirable incomes. Each species' umwelt must be taken into account when designing an experiment so that we can reproduce the most likely species-specific situations of a particular capacity to arise. In this case, presenting an ultimatum game to chimpanzees with the expectance of rejections, both in one-shot or in iterated games, does not seem the best option for revenge or functional punishment to emerge since chimpanzees' umwelt implies to act as a rational maximizer in the presence of available food. By contrast, the studies that have shown something akin to revenge in non-human primates have tested theft (Jensen et al., 2007b) and intragroup aggression (Aureli, Cozzolino, Cordischi, & Scucchi, 1992; de Waal & Luttrell, 1988), two definitely very chimpanzee common behaviors. Indeed, we have been able to reproduce punishment (indirect revenge) in the lab when agents tricked and took food away from chimpanzees (Studies 3 and 4). Some authors had already warned of the need to forget about human-look games and encouraged to look for more valid and comparable situations instead, at least when studying reciprocity: "The emphasis on these unique aspects of our behavior has sometimes distracted scientists from paying attention to the more common aspects of our daily lives, which share characteristics with those of our fellow primates. We invite friends for dinner, console others after a loss, intervene in ongoing fights, and even groom others. These small acts of altruism, which constitute a large part of our daily social life, tend to resemble those of nonhuman primates" (Schino & Aureli, 2010a).

As it can be seen, the beginning of this discussion exposes a criticism that not only applies to previous studies but also to ourselves and to the methodological choice we made to study functional punishment in Study 1. In Study 1 we wanted to see whether chimpanzee responders indirectly "taught" their partners what should be the best option to offer them, namely if functional punishment arose and worked out by changing the proposers' behavior. The problem was that we relied on rejections, and those never occurred. Therefore, without the possibility of punishment to happen, proposers' changes of behavior became also very unlikely. This initial criticism is a conclusion that we extract after conducting all the experiments of the present dissertation (obviously, if it had been a previous thought we had avoided conducting some of them). This conclusion has relevant implications in the evolutionary roots of punishment. Now that

we know that the essence of the ultimatum game (the risk of rejections) does not apply similarly to both species we can focus on discussing which the differential factors of human and chimpanzee responders are when playing MUG to gain more knowledge about the evolutionary pathway of punishment. Plus, now that we know that both young human and chimpanzee proposers are not consistently prosocial in a MUG, we can also look at their similarities. Finally, as we also know that both humans and chimpanzees are interested in seeing antisocial agents being punished we can also discuss about the motivational roots of punishment.

In Study 1 we found two interesting facts. The first is that chimpanzees were significantly prosocial (i.e. gave food) in no cost conditions in the DG. The second is that their prosocial offers decreased in the MUG. These results might seem difficult to be interpreted from a human perspective of how we understand and play the game. Rationally, we may think, if chimpanzees are rational maximizers, they should not be generous or, perhaps, not as generous in a DG as in a MUG. However, if we adopt the referred chimpanzees' umwelt, these two findings might make better sense. Being a rational maximizer and playing as a proposer imply not incurring costs (or incurring them at a minimum level in order to get more). Therefore being prosocial at no cost in repeated interactions in DG is an expected result: why not offering something to my partner if it is no costly for me? This does not necessarily mean that the chimpanzee proposer is expecting something in return. Positive reciprocity has been reported (grooming reciprocation: Schino & Aureli, 2010b) whereas short-term calculated (a synonym of intentional) reciprocity does not seem to consistently exist (Amici et al., 2014). Giving at no cost can just indicate some prosocial tendency (remember that Proctor et al, 2013, showed costly prosociality in their chimpanzee subjects). Prosocial tendencies at no cost are present in the wild: chimpanzees tolerate theft (Blurton Jones, 1984), which is similar to non-costly prosociality because generally the stolen food consists of peels or small pieces. Therefore what our DG indicates is that chimpanzees are prosocial preferentially when it is no costly, but that chimpanzee responders are not consistently reciprocating the partner's niceness. Note that this is already different from humans, who definitely incur costs when playing DG (Henrich et al., 2006). Why humans do incur costs but chimpanzees do not?

One key difference between humans and chimpanzees is the concept of group belonging present in humans. Group belonging, also referred as "group-mindedness" (Tomasello,

2014; Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012) is the collective perception of the individuals of a group. This group thinking allows the emergence of social and moral norms, such as equitable justice. Equitable justice is shown in societies which define themselves as a community of equals. If two players have the common perception to belong to a common abstract idea, such as the idea of equal society, both of them will act accordingly: proposers will offer something "fair" and responders will reject "unfair" outcomes. That is why human rejects lower offers from another human but not from a computer (Blount, 1995). Hence, if players do not have this sense of belonging with underlying abstract equality rules, no matter how many times we run DG or MUG: no individual will incur costs and no individual will reject non-zero outcomes.

The question here would be why things changed in the MUG. Contrary to the DG, in the MUG chimpanzees were each session less prosocial at no cost. Why did this happen? In our opinion we need again to adopt a chimpanzee perspective. Chimpanzee proposers offered predominantly selfishly in cost conditions. This meant offering 6/0 to the partner. Chimpanzee responders were also rational maximizers and therefore they accepted all non-zero offers but reject approximately half of zero-offers (rejecting zero outcomes is compatible with rational maximizers, see Henrich & Silk, 2013). As a proposer, facing 6/0 rejections could be interpreted as a warning to cooperate (ergo, functional punishment could arise). But it can be also interpreted with frustration for not obtaining the desirable outcome (McAuliffe et al., 2014; Roma, Silberberg, Ruggiero, & Suomi, 2006; Silberberg, Crescimbene, Addessi, Anderson, & Visalberghi, 2009). Frustration can lead to anger and this, in turn, can diminish the willing to cooperate. The ability to understand rejections as some social discomfort in the partner and change behavior accordingly is difficult. In fact, not even children are able to do that until the age of eight (Harbaugh et al., 2007). To the absence of a well-developed sense of group-mindedness that fueled with generosity initial offers, it might be added poor metacognition (low ability to think about how one is behaving; what consequences the behavior is producing and how that behavior can be changed to produce new better outcomes); some inability of full perspective taking (Cowell, Samek, List, & Decety, 2015; Sally & Hill, 2006) and potential high social comparison. In fact, another differential factor while playing MUG between chimpanzees and humans is social comparison, in other words, the fact that children showed disadvantageous IA.

Disadvantageous IA is a byproduct of group mindedness because it allows to comparing outcomes and protest when these are not equal (Raihani & Mcauliffe, 2017). Concretely, disadvantageous IA has proved to be very prominent in children, especially at 10 years of age.

The similarity between children and chimpanzees while playing MUG in Study 1 and Study 2 was that proposers shared their inability to anticipate what the partner will be willing to accept. In the first case, this inability did not allow cooperation or positive reciprocity to arise. In the second case, this inability (together with the existence of high social comparison in the responders) produced functional spite.

The interesting fact is that, regardless of the differences in the application of punishment (probably because abstract elements such as the concept of fairness might be only available for one species), both children and chimpanzees shared the motivation to watch punishment enacted. Concretely, they both enjoy significantly more watching those who previously and directly harmed them as opposed to those who were prosocial with them. Importantly, chimpanzees did not enjoy the same when the harm had not been directed towards them. A joyful experience of third-party punishment is attributable to some abstract thought such as "antisocial agents must be punished". The fact that chimpanzees do not show this kind of enjoyment is compatible with the absence of group mindedness that we referred earlier.

It is very likely that the antisocial action we chose with chimpanzees (i.e. a human tricking them with food and taking the solicited food away) was fundamental to elicit the motivation to watch punishment. Surely, if the harmed action was some previous bargaining, as the food had been given instead of stolen/taken away, the motivation would have not been present. Nevertheless, the very relevant thing is that punishment seems to be evolutionary desirable when the affected victim cannot execute it. This shows that the basis for functional spite is evolutionary common to humans and chimpanzees. The intriguing thing is whether this motivation (i.e. Schadenfreude) is a cause or a consequence for functional spite. At this point, we remain cautious to claim the ulterior or precursor nature of Schadenfreude.

In summary, we are confident to conclude that both species share some common grounds around punishment (such as the motivation to enjoy antisocial punishment), but the actual execution of punishment differs essentially because of their respective social

constructions (i.e. group mindedness), which have led to different reactions when facing the same scenarios (i.e. rational maximization vs. social comparison and disadvantageous IA).

Conclusiones y propuestas de futuro

Conclusions and future directions

Conclusiones y propuestas de futuro (castellano):

ESTUDIO 1: MINI-ULTIMÁTUM Y DICTATOR EN CHIMPANCÉS

1. Los chimpancés no juegan el MUG ni el DG de la misma forma que los humanos. Los humanos reivindican un determinado trato cuando juegan con otros humanos, mientras que los chimpancés se focalizan en las ganancis materiales que pueden tener.

2. Los chimpancés juegan de manera racional para maximizar, es decir, los que reciben no rechazan ofertas distintas de cero (el rechazo solo se produce la mitad de las veces que reciben una oferta igual a cero) y los que ofrecen lo hacen de forma generosa especialmente cuando no supone un coste para ellos. Es probable que la principal diferencia con los humanos sea la "conciencia de grupo/especie".

3. El rechazo no funciona como motor de cambio hacia una mayor generosidad. De hecho, los chimpancés que ofrecen disminuyen su generosidad tras los rechazos, posiblemente por la frustración que les supone no haber recibido comida.

4. Los chimpancés que ofrecen no parecen tener en cuenta la capacidad que tenga el compañero para rechazar, ya que no varían su generosidad entre juegos.

5. Los intentos comunicativos del chimpancé que recibe hacia el chimpancé que ofrece no tienen influencia sobre la generosidad de las ofertas.

6. No se da de manera consistente ni reciprocidad negativa ni reciprocidad positiva entre chimpancés.

7. Los juegos de ultimátum y dictator no parecen los más adecuados para explorar conceptos de castigo y justicia en chimpancés, ya que la metodología que siguen no está adaptada a las condiciones naturales y sociales de la especie.

Propuesta de futuro: se deberían plantear experimentos que reproduzcan conductas naturales en el laboratorio, a fin de poder juzgar con más precisión las conductas de cada especie. Por ejemplo, una acción "mala" podría ser quitar comida o pegar, mientras que una acción "buena" podría ser hacer grooming o intervenir en una pelea. Se deberían evitar las conductas de rechazo o distribución de comida, ya que no son típicas del repertorio conductual de los chimpancés.

ESTUDIO 2: MINI-ULTIMÁTUM EN NIÑOS

1. Hay indicios de venganza en niños de 6 y 10 años (rechazan en interacciones de un único ensayo).

2. Esta venganza parece determinada por la aversión a la inequidad, que es mayor en niños de mayor edad, posiblemente debido a la relevancia de la comparación social al inicio de la adolescencia.

3. La intención del compañero no parece tener una influencia en la respuesta de venganza en este tipo de juegos al menos hasta los 10 años. Es probable que sea detectada pero quede enmascarada por el elevado nivel de aversión a la inequidad.

4. Solo las niñas de 10 años son capaces de hacer ofertas al compañero en función de lo que a su edad es típicamente aceptable. Es probable que sea debido a la tendencia de las niñas por fijarse más en los aspectos sociales de una interacción antes que en las ganancias materiales.

5. La metodología influye significativamente en las respuestas de venganza. El mismo paradigma de juego (ultimátum) con pequeños cambios metodológicos produce resultados muy diferentes.

Propuesta de futuro: próximas investigaciones deberían reproducir escenarios reales en el laboratorio (por ejemplo, seguir el direct method antes que el strategy method). Esto supone en ocasiones un desembolso de tiempo y recursos económicos mayor, pero asegura la validez de los resultados. Además, futuros estudios deberían tener en cuenta la incidencia de la aversión a la inequidad y la comparación social para diseñar condiciones de control que las excluyeran, si fuera necesario para sus objetivos de investigación. Sería de hecho interesante explorar cuándo desaparece en la ontogenia esa comparación social y a qué razones se debe.

ESTUDIOS 3 Y 4: SCHADENFREUDE EN NIÑOS Y CHIMPANCÉS

1. Humanos (a partir de 6 años) y chimpancés comparten interés por contemplar cómo recibe un castigo alguien que les infligió un daño previo.

2. Es probable que el concepto de moralidad como el imperativo de que cualquier evento negativo debe seguirse de un castigo parece propio de los humanos, ya que los

chimpancés no incurrieron gastos para ver cómo recibía un castigo un agente que había dañado a otro de su misma especie. Esto puede estar provocado, como se señalaba en las conclusiones del estudio 1, por la posible ausencia de "concepto de grupo/especie" en los chimpancés.

Propuesta de futuro: sería muy interesante conocer en más profundidad las reacciones emocionales de los chimpancés. En humanos contamos con indicadores más o menos universales que se asocian a determinadas emociones (eg. sonrisa-alegría, llanto-tristeza), aunque se ha visto en este estudio que en ocasiones se produce la combinación de algunos opuestos y genera confusión en la interpretación (eg. fruncir el ceño y sonreír al mismo tiempo). En chimpancés, en cambio, seguimos empleando la clasificación de Goodall sobre las vocalizaciones sin tener un conocimiento claro sobre cuándo se producen y qué pueden significar para los miembros de esa especie. Futuros estudios podrían intentar profundizar en la búsqueda de asociaciones consistentes entre gestos comunicativos y emociones, especialmente en el ámbito de la moralidad.

Conclusions and future directions (English version):

STUDY 1: MINI-ULTIMATUM AND DICTATOR IN CHIMPANZEES

1. Chimpanzees do not play MUG or DG in the same way as humans do. Humans claim to be treated in a especial way when they play with other humans, while chimpanzees focus on the material gains they can obtain.

2. Chimpanzees play as rational maximizers do, that is, recipients do not reject offers different from zero (rejections occur half of the time facing zero-offers) and proposers are generous especially when it is no costly. It is likely that the main difference between both species is the "group mididness" that humans have.

3. Rejection does not promote generosity. In fact, chimpanzee proposers offer less prosocially after rejections, possibly due to frustration after not receiving food.

4. Chimpanzees proposers do not seem to take into account their partner's capacity to reject, since their generosity between games does not vary.

5. Communicative attempts have no influence on the proposer's generosity.

6. There is no consistent negative or positive reciprocity between chimpanzees.

7. The ultimatum and dictator games do not seem to be the most appropriate way to explore punishment and justice in chimpanzees, since the methdology is not adapted to the natural and social conditions of the species.

Future directions: there is the challenge to design experiments that reproduce natural behaviors in the laboratory in order to be able to judge more accurately the conuctas of each species. For example, a "bad" action could be to remove food or paste, while a "good" action could be to do grooming or intervene in a fight. Rejection or food distribution behaviors should be avoided, as they are not typical of the behavioral repertoire of chimpanzees.

STUDY 2: MINI-ULTIMATUM IN CHILDREN

1. There are indications of revenge in children 6 and 10 years old (reject in single-trial interactions).

2. This revenge seems determined by inequity aversion, which is higher in older children, possibly due to the relevance of social comparison in early adolescence.

3. The proposer's intention does not seem to have an influence on the revenge response in this type of games at least until 10 years of age. Children probably detect intention but it is masked by the high level of inequity aversion.

4. Only 10-year-old girls are able to offer based on what is typically acceptable at their age. This capacity is probably developed earlier in girls because they tend to focus more on the social aspects of an interaction rather than on the material gains.

5. The methodology employed significantly influences revenge responses. The same paradigm (ultimatum game) with few methodological changes produces very different results.

Future directions: future research should reproduce real scenarios in the laboratory (for example, conducting the direct method rather than the strategy method). This sometimes implies a greater expenditure of time and resources, but it ensures the validity of the results. In addition, future studies should take into account the prevalence of inequity aversion and social comparison in children in order to design appropriate control conditions, if necessary. It would be interesting to explore when social comparison vanishes across ontogeny and under which reasons.

STUDIES 3 AND 4: SCHADENFREUDE IN CHILDREN AND CHIMPANZEES

1. Humans (from 6 years old) and chimpanzees share a common interest to contemplate how a person that inflicted a previous damage receives some punishment.

2. Morality understood as the imperative of "any negative event must be followed by a punishment" is probably uniquely human, since chimpanzees did not incur costs to continue watching the punishment of an agent who had previously harmed a conspecific. As noted in the conclusions of study 1, the absence of third-party concern might be potentially due to the absence of "group mindedness" in chimpanzees.

Future directions: it would be very interesting to know in depth the emotional reactions of chimpanzees. Humans have universal indicators associated with certain emotions (eg. smile-happiness, cry-sadness), although sometimes the combination of some opposite indicators occurs leading to confusion (eg. frown and smiles co-

occurring in Study 3). In chimpanzees, however, we continue using Goodall's classification of vocalizations without having a clear understanding of wha they especifically mean for the members of that species and when they occur. Future research may try to look for consistent associations between communicative gestures and emotions in chimpanzees, especially in the realm of morality.

Table Index:

Figure Index:

References:

Amici, F., Aureli, F., Mundry, R., Amaro, A. S., Barroso, A. M., Ferretti, J., & Call, J. (2014). Calculated reciprocity? A comparative test with six primate species. *Primates*, *55*(3), 447–457. http://doi.org/10.1007/s10329-014-0424-4

Anderson, J. R., Kuroshima, H., Takimoto, A., Fujita, K., Vaish, A., Carpenter, M., … Bshary, R. (2013). Third-party social evaluation of humans by monkeys. *Nature Communications*, *4*, 1561. http://doi.org/10.1038/ncomms2495

Anderson, J. R., Takimoto, A., Kuroshima, H., & Fujita, K. (2013). Capuchin monkeys judge third-party reciprocity. *Cognition*, *127*(1), 140–146. http://doi.org/10.1016/j.cognition.2012.12.007

Anonymous. (1998). To Reveal or Not to Reveal: A Theoretical Model of Anonymous Communication. *Communication Theory*, *8*(4), 381–407. http://doi.org/10.1111/j.1468-2885.1998.tb00226.x

Arenal, C. (1867). *El reo, el pueblo y el verdugo o la ejecución pública de la pena de muerte*. Madrid: Tipografía de Estrada, Díaz y López.

Aureli, F., Cozzolino, R., Cordischi, C., & Scucchi, S. (1992). Kin-oriented redirection among Japanese macaques: an expression of a revenge system? *Animal Behaviour*, *44*, 283–291. http://doi.org/10.1016/0003-3472(92)90034-7

Axelrod, R., & Hamilton, W. D. (1981). The Evolution of Cooperation. *Source: Science, New Series*, *211*(4489), 1390–1396. Retrieved from http://www.jstor.org/stable/1685895

Baayen, R. H. (2008). Analyzing linguistic data: A practical introduction to statistics using R. *Processing*, *2*(3), 353. http://doi.org/10.1558/sols.v2i3.471

Barash, D. P., & Lipton, J. E. (2011). *Payback: why we retaliate, redirect aggression and seek revenge*. Oxford: Oxford University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. http://doi.org/10.1016/j.jml.2012.11.001

Barreca, R. (1995). *Sweet Revenge (The Wicked Delights of Getting Even)*. New York: Harmony Books.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). http://doi.org/10.18637/jss.v067.i01

Batson, C. D. (1991). *The Altruism Question: Toward a Social-Psychological Answer*. Hillsdale, NJ: Erlbaum.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is Empathic Emotion a Source of Altruistic Motivation? *Journal of Personality and Social Psychology*, *40*(2), 290–302.

Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children's altruistic behavior in the

dictator game. *Evolution and Human Behavior*, *28*(3), 168–175. http://doi.org/10.1016/j.evolhumbehav.2006.10.003

Beran, M. J., & Evans, T. A. (2009). Delay of gratification by chimpanzees (Pan troglodytes) in working and waiting situations. *Behavioural Processes*, *80*(2), 177–181. http://doi.org/10.1016/j.beproc.2008.11.008

Beutel, A. M., & Johnson, M. K. (2004). Gender and prosocial values during adolescence: A Research Note. *The Sociological Quarterly*, *45*(2), 379–393. http://doi.org/10.1111/j.1533-8525.2004.tb00017.x

Blake, P. R., & McAuliffe, K. (2011). "I had so much it didn't seem fair": Eight-year-olds reject two forms of inequity. *Cognition*, *120*(2), 215–224. http://doi.org/10.1016/j.cognition.2011.04.006

Blake, P. R., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: the knowledge-behavior gap. *Trends in Cognitive Sciences*, *18*(11), 559–61. http://doi.org/10.1016/j.tics.2014.08.003

Blake, P. R., Piovesan, M., Montinari, N., Warneken, F., & Gino, F. (2015). Prosocial norms in the classroom: The role of self-regulation in following norms of giving. *Journal of Economic Behavior & Organization*, *115*, 18–29. http://doi.org/10.1016/j.jebo.2014.10.004

Blount, S. (1995). When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences. *Organizational Behavior and Human Decision Processes*, *63*(2), 131–144. http://doi.org/10.1006/obhd.1995.1068

Blurton Jones, N. G. (1984). A selfish origin for human food sharing: Tolerated theft. *Ethology and Sociobiology*, *5*(1), 1–3. http://doi.org/10.1016/0162-3095(84)90030-X

Bowden, M. (2012). *The finish: The killing of Osama bin Laden*. London: Atlantic Books.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(6), 3531–5. http://doi.org/10.1073/pnas.0630443100

Brosnan, S. F. (2011). A hypothesis of the co-evolution of cooperation and responses to inequity. *Frontiers in Neuroscience*, *5*, 43. http://doi.org/10.3389/fnins.2011.00043

Brosnan, S. F. (2012). Introduction to "Justice in Animals." *Social Justice Research*, *25*(2), 109–121. http://doi.org/10.1007/s11211-012-0156-9

Brosnan, S. F. (2013). Justice- and fairness-related behaviors in nonhuman primates. *Proceedings of the National Academy of Sciences of the United States of America*, *110 Suppl*, 10416–23.

Brosnan, S. F., & de Waal, F. B. (2014). Evolution of responses to (un)fairness. *Science*, *346*(6207), 1251776. http://doi.org/10.1126/science.1251776

Brosnan, S. F., Salwiczek, L., & Bshary, R. (2010). The interplay of cognition and cooperation. *Philosophical Transactions of the Royal Society of London. Series B,*

Biological Sciences, *365*(1553), 2699–710. http://doi.org/10.1098/rstb.2010.0154

Bueno-Guerra, N. (2012). Vengeance with a vengence: is it worth to study it? *Anuario de Psicología Jurídica*, *22*, 95–110.

Bueno-Guerra, N., Leiva, D., Colell, M., & Call, J. (2016). Do sex and age affect strategic behavior and inequity aversion in children? *Journal of Experimental Child Psychology*, *150*, 285–300. http://doi.org/10.1016/j.jecp.2016.05.011

Bullinger, A. F., Burkart, J. M., Melis, A. P., & Tomasello, M. (2013). Bonobos, Pan paniscus, chimpanzees, Pan troglodytes, and marmosets, Callithrix jacchus, prefer to feed alone. *Animal Behaviour*, *85*(1), 51–60. http://doi.org/10.1016/j.anbehav.2012.10.006

Butterfield, K. D., Trevino, L. K., & Ball, G. A. (1996). Punishment from the manager's perspective: A grounded investigation and inductive model. *Academy of Management Journal*, *39*(6), 1479–1512. http://doi.org/10.2307/257066

Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). "Unwilling" versus "unable": chimpanzees' understanding of human intentional action. *Developmental Science*, *7*(4), 488–498. http://doi.org/10.1111/j.1467-7687.2004.00368.x

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

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why do we punish?: Deterrence and just deserts as motives for punishment. *Journal of Personality and Social Psychology*, *83*(2), 284–299. http://doi.org/10.1037/0022-3514.83.2.284

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

Charness, G., & Gneezy, U. (2008). What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization*, *68*, 29–35. http://doi.org/10.1016/j.jebo.2008.03.001

Chernyak, N., & Sobel, D. M. (2016). "But he didn't mean to do it": Preschoolers correct punishments imposed on accidental transgressors. *Cognitive Development*, *39*, 13–20. http://doi.org/10.1016/j.cogdev.2016.03.002

Cikara, M., & Fiske, S. T. (2013). Their pain, our pleasure: stereotype content and schadenfreude. *Annals of the New York Academy of Sciences*, *1299*(1), 52–59. http://doi.org/10.1111/nyas.12179

Clutton-Brock, T. H., & Parker, G. A. (1995). Punishment in animal societies. *Nature*, *373*(6511), 209–216. http://doi.org/10.1038/373209a0

Cooper, D. J., & Dutcher, E. G. (2011). The dynamics of responder behavior in ultimatum games: a meta-study. *Experimental Economics*, *14*(4), 519–546. http://doi.org/10.1007/s10683-011-9280-x

Cota-McKinley, A. L., Woody, W. D., & Bell, P. A. (2001). Vengeance: Effects of gender, age, and religious background. *Aggressive Behavior*, *27*(5), 343–350.

http://doi.org/10.1002/ab.1019

Cowell, J. M., Samek, A., List, J., & Decety, J. (2015). The curious relation between theory of mind and sharing in preschool age children. *PLoS ONE*, *10*(2). http://doi.org/10.1371/journal.pone.0117947

Crombag, H., Rassin, E., & Horselenberg, R. (2003). ON VENGEANCE. *Psychology, Crime & Law*, *9*(4), 333–344. http://doi.org/10.1080/1068316031000068647

Damon, W. (1977). *The social world of the child*. San Francisco: Jossey-Bass.

Daniels, M. (1969). Pathological vindictiveness and the vindictive character. *Psychoanalytic Review*, *56*(2), 169–96.

Darwin, C. (1871). *The Descent of Man*. London: Murray.

de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The Neural Basis of Altruistic Punishment. *Science*, *305*(5688).

de Waal, F. B. M., & Brosnan, S. F. (2006). Simple and complex reciprocity in primates. In *Cooperation in Primates and Humans* (pp. 85–105). Berlin, Heidelberg: Springer Berlin Heidelberg.

de Waal, F. B. M. (1996). *Good natured : the origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.

de Waal, F. B. M., & Luttrell, L. M. (1988). Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology*, *9*(2), 101–118. http://doi.org/10.1016/0162-3095(88)90016-7

Dobson, A. J., & Barnett, A. G. (2008). *An introduction to generalized linear models*. CRC Press.

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

Eckel, C. C., & Grossman, P. J. (2008). Chapter 57 Differences in the Economic Decisions of Men and Women: Experimental Evidence. In *Handbook of Experimental Economics Results* (Vol. 1, pp. 509–519). http://doi.org/10.1016/S1574-0722(07)00057-1

Eckel, C., & Grossman, P. (2001). Chivalry and solidarity in ultimatum games. *Economic Inquiry*, *39*(2), 171–188. http://doi.org/10.1111/j.1465-7295.2001.tb00059.x

Eisenberg, N., Fabes, R. A., Karbon, M., Murphy, B. C., Wosinski, M., Polazzi, L., … Juhnke, C. (1996). The relations of children's dispositional prosocial behavior to emotionality, regulation, and social functioning. *Child Development*, *67*(September), 974–992. http://doi.org/10.2307/1131874

Eisenberg, N., Fabes, R. A., Murphy, B., Karbon, M., Maszk, P., Smith, M., … Suh, K. (1994). The relations of emotionality and regulation to dispositional and situational

empathy-related responding. *Journal of Personality and Social Psychology*, *66*(4), 776–97.

Ellsworth, P. C., & Gross, S. R. (1994). Hardening of the Attitudes: Americans' Views on the Death Penalty. *Journal of Social Issues*, *50*(2), 19–52. http://doi.org/10.1111/j.1540-4560.1994.tb02409.x

Elster, J. (1990). Norms of Revenge. *Ethics*, *100*(4), 862–885. http://doi.org/10.1086/293238

Engelmann, J. M., Over, H., Herrmann, E., & Tomasello, M. (2013). Young children care more about their reputation with ingroup members and potential reciprocators. *Developmental Science*, *16*(6), 952–958. http://doi.org/10.1111/desc.12086

Fabes, R. a., Eisenberg, N., & Eisenbud, L. (1993). Behavioral and physiological correlates of children's reactions to others in distress. *Developmental Psychology*, *29*(4), 655–663. http://doi.org/10.1037/0012-1649.29.4.655

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the Nature of Fair Behavior. *Economic Inquiry*, *41*(1), 20–26. http://doi.org/10.1093/ei/41.1.20

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., & Nairn, K. (2005). Resentment, envy, schadenfreude, and sympathy: Effects of own and other's deserved or undeserved status. *Australian Journal of Psychology*, *57*(2), 87–102. http://doi.org/10.1080/00049530500048672

Feather, N. T. (2006). Deservingness and emotions: Applying the structural model of deservingness to the analysis of affective reactions to outcomes. *European Review of Social Psychology*, *17*(1), 38–73. http://doi.org/10.1080/10463280600662321

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

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–91. http://doi.org/10.1038/nature02043

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. http://doi.org/10.1016/S1090-5138(04)00005-4

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, *13*(1), 1–25. http://doi.org/10.1007/s12110-002-1012-7

Fehr, E., & Gächter, S. (2000a). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994. http://doi.org/10.1126/science.151.3712.867-a

Fehr, E., & Gächter, S. (2000b). Fairness and Retaliation: The Economics of Reciprocity. *Journal of Economic Perspectives*, *14*(3), 159–182. http://doi.org/10.1257/jep.14.3.159

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140. http://doi.org/10.1038/415137a

Flack, J. C., Girvan, M., de Waal, F. B. M., & Krakauer, D. C. (2006). Policing stabilizes construction of social niches in primates. *Nature*, *439*(7075), 426–429. http://doi.org/10.1038/nature04326

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, *65*(1), 47–55. http://doi.org/10.1007/s00265-010-1038-5

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

Funk, F., McGeer, V., & Gollwitzer, M. (2014). Get the Message: Punishment Is Satisfying If the Transgressor Responds to Its Communicative Intent. *Personality and Social Psychology Bulletin*, *40*(8), 986–997. http://doi.org/10.1177/0146167214533130

García-Gallego, A., Georgantzís, N., & Jaramillo-Gutiérrez, A. (2012). Gender differences in ultimatum games: Despite rather than due to risk attitudes. *Journal of Economic Behavior & Organization*, *83*(1), 42–49. http://doi.org/10.1016/j.jebo.2011.06.012

Gardner, A., & West, S. A. (2006). Spite. *Current Biology : CB*, *16*(17), R662-4. http://doi.org/10.1016/j.cub.2006.08.015

Gilby, I. C. (2006). Meat sharing among the Gombe chimpanzees: harassment and reciprocal exchange. *Animal Behaviour*, *71*(4), 953–963. http://doi.org/10.1016/j.anbehav.2005.09.009

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, *206*(2), 169–79. http://doi.org/10.1006/jtbi.2000.2111

Gollwitzer, M., Meder, M., & Schmitt, M. (2011). What gives victims satisfaction when they seek revenge? *European Journal of Social Psychology*, *41*(3), 364–374. http://doi.org/10.1002/ejsp.782

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge, MA: Belknap Press of Harvard University Press.

Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, *35*(1), 1–15. http://doi.org/10.1017/S0140525X11000069

Gummerum, M., & Chu, M. T. (2014). Outcomes and intentions in children's, adolescents', and adults' second- and third-party punishment behavior. *Cognition*, *133*(1), 97–103. http://doi.org/10.1016/j.cognition.2014.06.001

Güroğlu, B., van den Bos, W., & Crone, E. A. (2009). Fairness considerations: increasing understanding of intentionality during adolescence. *Journal of Experimental Child Psychology*, *104*(4), 398–409. http://doi.org/10.1016/j.jecp.2009.07.002

Güth, W., & Kocher, M. G. (2014). More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior & Organization*, *108*, 396–409. http://doi.org/10.1016/j.jebo.2014.06.006

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388. http://doi.org/10.1016/0167-2681(82)90011-7

Hamilton, W. D. (1964). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, *7*(1), 17–52. http://doi.org/10.1016/0022-5193(64)90039-6

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*(1), 30–39. http://doi.org/10.1016/j.cogdev.2010.09.001

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559. http://doi.org/10.1038/nature06288

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 19931–6. http://doi.org/10.1073/pnas.1110306108

Hanus, D., & Call, J. (2007). Discrete quantity judgments in the great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus): The effect of presenting whole sets versus item-by-item. *Journal of Comparative Psychology*, *121*(3), 241–249. http://doi.org/10.1037/0735-7036.121.3.241

Hanus, D., Mendes, N., Tennie, C., & Call, J. (2011). Comparing the performances of apes (gorilla gorilla, pan troglodytes, pongo pygmaeus) and human children (homo sapiens) in the floating peanut task. *PLoS ONE*, *6*(6). http://doi.org/10.1371/journal.pone.0019555

Harbaugh, W. T., Krause, K. S., & Liday, S. J. (2003). Bargaining by Children. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.436504

Harbaugh, W. T., Krause, K., & Vesterlund, L. (2007). Learning to bargain. *Journal of Economic Psychology*, *28*(1), 127–142. http://doi.org/10.1016/j.joep.2006.03.005

Hauser, M. D. (1992). Costs of deception: cheaters are punished in rhesus monkeys (Macaca mulatta). *Proceedings of the National Academy of Sciences of the United States of America*, *89*(24), 12137–12139. http://doi.org/10.1073/pnas.89.24.12137

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley & Sons.

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural Responses to Ingroup and Outgroup Members' Suffering Predict Individual Differences in Costly Helping. *Neuron*, *68*(1), 149–160. http://doi.org/10.1016/j.neuron.2010.09.003

Helwig, C. C., Zelazo, P. D., & Wilson, M. (2001). Children's judgements of psychological harm in normal and noncanonical situations. *Child Development*, *72*(1), 66–81.

Henrich, J., & Boyd, R. (2001). Why people punish defectors - Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal Of Theoretical Biology*, *208*(1), 79–89. http://doi.org/10.1006/jtbi.2000.2202

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., … Ziker, J. (2006). Costly Punishment Across Human Societies. *Science*, *312*(5781).

Henrich, J., & Silk, J. B. (2013). Interpretative problems with chimpanzee ultimatum game. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(33), E3049. http://doi.org/10.1073/pnas.1307007110

Henrich, J., & Silk, J. B. (2013). Supplemental to: Interpretative Problems with Chimpanzee Ultimatum Games. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2250049

Herrmann, E., Keupp, S., Vaish, A., Tomasello, M., Moumbaka, A., Sitou, M., … Musumba, G. (2012). Direct and Indirect Reputation Formation in Nonhuman Great Apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus) and Human Children (Homo sapiens). http://doi.org/10.1037/a0028929

Hill, K., Barton, M., & Hurtado, A. M. (2009). The emergence of human uniqueness: Characters underlying behavioral modernity. *Evolutionary Anthropology: Issues, News, and Reviews*, *18*(5), 187–200. http://doi.org/10.1002/evan.20224

Ho, R., ForsterLee, L., ForsterLee, R., & Crofts, N. (2002). Justice versus vengeance: motives underlying punitive judgements. *Personality and Individual Differences*, *33*(3), 365–377. http://doi.org/10.1016/S0191-8869(01)00161-1

Hoffman, E., & Spitzer, .L. (1985). Entitlements, Rights, and Fairness: An Experimental Examination of Subjects' Concepts of Distributive Justice on JSTOR. *The Journal of Legal Studies*, *14*(2), 259–297.

Honderich, T. (1969). *Punishment - The supposed justifications*. London: Pluto Press.

Horney, K. (1948). The value of vindictiveness. *The American Journal of Psychoanalysis*, *8*(1), 3–12. http://doi.org/10.1007/BF01871591

House, B. R., Henrich, J., Brosnan, S. F., & Silk, J. B. (2012). The ontogeny of human prosociality: behavioral experiments with children aged 3 to 8. *Evolution and Human Behavior*, *33*(4), 291–308. http://doi.org/10.1016/j.evolhumbehav.2011.10.007

Jacoby, S. (1983). *Wild Justice: The Evolution of Revenge*. New York: Harper & Row.

Jaffee, S., & Hyde, J. S. (2000). Gender differences in moral orientation: A meta-analysis. *Psychological Bulletin*, *126*(5), 703–726. http://doi.org/http://dx.doi.org/10.1037/0033- 2909.126.5.703

Jensen, K. (2010). Punishment and spite, the dark side of cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2635–2650. http://doi.org/10.1098/rstb.2010.0146

Jensen, K., Call, J., & Tomasello, M. (2007a). Chimpanzees are rational maximizers in

an ultimatum game. *Science (New York, N.Y.)*, *318*(5847), 107–9. http://doi.org/10.1126/science.1145850

Jensen, K., Call, J., & Tomasello, M. (2007b). Chimpanzees are vengeful but not spiteful. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(32), 13046–50. http://doi.org/10.1073/pnas.0705555104

Jensen, K., Call, J., & Tomasello, M. (2013). Chimpanzee responders still behave like rational maximizers. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(20), E1837. http://doi.org/10.1073/pnas.1303627110

Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(35), 12710–5. http://doi.org/10.1073/pnas.1402280111

Kaiser, I., Jensen, K., Call, J., Tomasello, M., Fehr, E., Fischbacher, U., … Yamagishi, T. (2012). Theft in an ultimatum game: chimpanzees and bonobos are insensitive to unfairness. *Biology Letters*, *8*(6), 942–5. http://doi.org/10.1098/rsbl.2012.0519

Kaufman, W. R. P. (2013). *Honor and Revenge: A Theory of Punishment*. New York: Springer.

Kawai, N., Yasue, M., Banno, T., Ichinohe, N., Nowak, M., Highfield, R., … Brown, V. (2014). Marmoset monkeys evaluate third-party reciprocity. *Biology Letters*, *10*(5), 20140058. http://doi.org/10.1098/rsbl.2014.0058

Keysar, B., Converse, B. a, Wang, J., & Epley, N. (2008). Reciprocity is not give and take. *Psychological Science*, *19*(12), 1280–1286. http://doi.org/10.1111/j.1467-9280.2008.02223.x

Kurzweil, A. (2015). *Whipping Boy. The forty-year search for my twelve-year-old bully*. New York: HarperCollins.

Lambert, A. J., Peak, S. A., Eadeh, F. R., & Schott, J. P. (2014). How do you feel now? On the perceptual distortion of extremely recent changes in anger. *Journal of Experimental Social Psychology*, *52*, 82–95. http://doi.org/10.1016/j.jesp.2014.01.004

Lambert, A. J., Scherer, L. D., Schott, J. P., Olson, K. R., Andrews, R. K., O'Brien, T. C., & Zisser, A. R. (2010). Rally effects, threat, and attitude change: An integrative approach to understanding the role of emotion. *Journal of Personality and Social Psychology*, *98*(6), 886–903. http://doi.org/10.1037/a0019086

Leimar, O., & Hammerstein, P. (2010). Cooperation for direct fitness benefits. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1553), 2619–2626. http://doi.org/10.1098/rstb.2010.0116

Leman, P. J., Keller, M., Takezawa, M., & Gummerum, M. (2009). Children's and Adolescents' Decisions about Sharing Money with Others. *Social Development*, *18*(3), 711–727. http://doi.org/10.1111/j.1467-9507.2008.00486.x

Lihoreau, T. (2011). *Schadenfreude: The Little Book of Black Delights*. London: Elliott

& Thompson, LTD.

Lin, H., & Sunder, S. (2001). Using Experimental Data to Model Bargaining Behavior in Ultimatum Games. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.268755

Liu, B., Huang, Z., Xu, G., Jin, Y., Chen, Y., Li, X., … Jing, J. (2016). Altruistic sharing behavior in children: Role of theory of mind and inhibitory control. *Journal of Experimental Child Psychology*, *141*, 222-228. http://doi.org/10.1016/j.jecp.2015.09.010

Lucea Ayala, V. (2004). Reos, verdugos y muchedumbres: la percepción popular de la penalidad y la pena de muerte. Zaragoza. 1855-1915. *Revista de Historia Jerónimo Zurita*, *76–77*, 129–158.

Maccoby, E. E. (2002). Gender and Group Process: A Developmental Perspective. *Current Directions in Psychological Science*, *11*(2), 54–58. http://doi.org/10.1111/1467-8721.00167

Madeira, J. L. (2012). *Killing McVeigh : the death penalty and the myth of closure*. New York University Press.

Marchetti, A., Castelli, I., Harlé, K. M., & Sanfey, A. G. (2011). Expectations and outcome: The role of Proposer features in the Ultimatum Game. *Journal of Economic Psychology*, *32*(3), 446–449. http://doi.org/10.1016/j.joep.2011.03.009

Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., … Tracer, D. (2008). More "altruistic" punishment in larger societies. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*(1634).

Matsusaka, T. (2004). When does play panting occur during social play in wild chimpanzees? *Primates*, *45*(4), 221–229. http://doi.org/10.1007/s10329-004-0090-z

McAuliffe, K., Blake, P. R., & Warneken, F. (2014). Children reject inequity out of spite. *Biology Letters*, *10*(12), 20140743–20140743. http://doi.org/10.1098/rsbl.2014.0743

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, *134*, 1–10. http://doi.org/10.1016/j.cognition.2014.08.013

McCullough, M. E. (2008). *Beyond revenge : the evolution of the forgiveness instinct*. San Francisco, CA: Jossey-Bass.

McCullough, M. E., Kurzban, R., & Tabak, B. A. (2012). Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences*, *36*(1), 1–15. http://doi.org/10.1017/S0140525X11002160

McNamee, M. (2008). *Sports, Virtues and Vices: Morality Plays*. New York: Routledge.

Melis, A. P., & Semmann, D. (2010). How is human cooperation different? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1553), 2663–74. http://doi.org/10.1098/rstb.2010.0157

Mitford, A. B. (1871). The forty-seven ronins. In *Tales of Old Japan. Folklore, Fairytales, Ghost Stories and Legends of the Samurai* (pp. 1–20). London: Macmillan and Company.

Montes Luengo, J. (1897). *La pena de muerte y el derecho de indulto*. Madrid: Imprenta de López Aguado.

Moore, A., & Lloyd, D. (1990). *V for Vendetta*. New York: D Comics (Vertigo).

Murnighan, J. K., & Saxon, M. S. (1998). Ultimatum bargaining by children and adults. *Journal of Economic Psychology*, *19*(4), 415–445. http://doi.org/10.1016/S0167-4870(98)00017-8

Murphy, J. G. (2011). *Getting Even: Forgiveness and Its Limits*. http://doi.org/10.1093/acprof:oso/9780195178555.001.0001

Neely, M. N., Walter, E., Black, J. M., & Reiss, A. L. (2012). Neural Correlates of Humor Detection and Appreciation in Children. *Journal of Neuroscience*, *32*(5), 1784–1790. http://doi.org/10.1523/JNEUROSCI.4172-11.2012

Nishida, T., Zamma, K., Matsusaka, T., Inaba, A., & McGrew, W. C. (2010). *Chimpanzee behavior in the wild: an audio-visual encyclopedia*. Springer Science & Business Media.

Noë, R., & Hammerstein, P. (1994). Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral Ecology and Sociobiology*, *35*(1), 1–11. http://doi.org/10.1007/BF00167053

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge, UK: Cambridge University Press.

Overgaauw, S., Güroğlu, B., & Crone, E. A. (2012). Fairness Considerations When I Know More than You Do: Developmental Comparisons. *Frontiers in Psychology*, *3*, 424. http://doi.org/10.3389/fpsyg.2012.00424

Persky, J. (1995). Retrospectives: the ethology of homo economicus. *The Journal of Economic Perspectives*, *9*(2), 221–231. http://doi.org/10.2307/2138175

Proctor, D., Williamson, R. A., de Waal, F. B. M., & Brosnan, S. F. (2013a). Chimpanzees play the ultimatum game. *Proceedings of the National Academy of Sciences*, *110*(6), 2070–2075. http://doi.org/10.1073/pnas.1220806110

Proctor, D., Williamson, R. A., de Waal, F. B. M., & Brosnan, S. F. (2013b). Reply to Jensen et al.: Equitable offers are not rationally maximizing. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(20), E1838.

R Core Team. (2016). R: A language and environment for statistical computing. Viena: Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Raihani, N. J., & Mcauliffe, K. (2017). Human punishment is motivated by inequity aversion, not a desire for reciprocity. http://doi.org/10.1098/rsbl.2012.0470

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's Dilemma*. Michigan: The

Unviersity of Michigan Press.

Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No third-party punishment in chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(37), 14824–9. http://doi.org/10.1073/pnas.1203179109

Rochat, P., Dias, M. D. G., Guo Liping, G., Broesch, T., Passos-Ferreira, C., Winning, A., & Berg, B. (2009). Fairness in Distributive Justice by 3- and 5-Year-Olds Across Seven Cultures. *Journal of Cross-Cultural Psychology*, *40*(3), 416–442. http://doi.org/10.1177/0022022109332844

Roma, P. G., Silberberg, A., Ruggiero, A. M., & Suomi, S. J. (2006). Capuchin monkeys, inequity aversion, and the frustration effect. *Journal of Comparative Psychology*, *120*(1), 67–73. http://doi.org/10.1037/0735-7036.120.1.67

Rothbart, M. K., Sheese, B. E., Rueda, M. R., & Posner, M. I. (2011). Developing Mechanisms of Self-Regulation in Early Life. *Emotion Review : Journal of the International Society for Research on Emotion*, *3*(2), 207–213. http://doi.org/10.1177/1754073910387943

Russell, J. B., & Alexander, B. (2007). *A history of witchcraft, sorcerers, heretics & pagans*. USA: Thames & Hudson.

Russell, Y. I., Call, J., & Dunbar, R. I. M. (2008). Image scoring in great apes. *Behavioural Processes*, *78*(1), 108–111. http://doi.org/10.1016/j.beproc.2007.10.009

Saad, G., & Gill, T. (2001). Sex Differences in the Ultimatum Game: An Evolutionary Psychology Perspective. *Journal of Bioeconomics*, *3*(2/3), 171–193. http://doi.org/10.1023/A:1020583425623

Sally, D., & Hill, E. (2006). The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. *Journal of Economic Psychology*, *27*, 73–97. http://doi.org/10.1016/j.joep.2005.06.015

Samson, A. C., Hempelmann, C. F., Huber, O., & Zysset, S. (2009). Neural substrates of incongruity-resolution and nonsense humor. *Neuropsychologia*, *47*(4), 1023–1033. http://doi.org/10.1016/j.neuropsychologia.2008.10.028

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology : Official Journal of the International Society for Behavioral Ecology*, *20*(2), 416–420. http://doi.org/10.1093/beheco/arn145

Schino, G., & Aureli, F. (2010a). Primate reciprocity and its cognitive requirements. *Evolutionary Anthropology: Issues, News, and Reviews*, *19*(4), 130–135. http://doi.org/10.1002/evan.20270

Schino, G., & Aureli, F. (2010b). The relative roles of kinship and reciprocity in explaining primate altruism. *Ecology Letters*, *13*(1), 45–50. http://doi.org/10.1111/j.1461-0248.2009.01396.x

Schmidt, M. F. H., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition*,

124(3), 325–33. http://doi.org/10.1016/j.cognition.2012.06.004

Schoek, H. (1987). *Envy*. New York: Liberty Fund.

Schulz, K., Rudolph, A., Tscharaktschiew, N., & Rudolph, U. (2013). Daniel has fallen into a muddy puddle - Schadenfreude or sympathy? *British Journal of Developmental Psychology*, *31*(4), 363–378. http://doi.org/10.1111/bjdp.12013

Schumann, K., & Ross, M. (2010). The Benefits, Costs, and Paradox of Revenge. *Social and Personality Psychology Compass*, *4*(12), 1193–1205. http://doi.org/10.1111/j.1751-9004.2010.00322.x

Shamay-Tsoory, S. G., Ahronberg-Kirschenbaum, D., & Bauminger-Zviely, N. (2014). There is no joy like malicious joy: Schadenfreude in young children. *PLoS ONE*, *9*(7). http://doi.org/10.1371/journal.pone.0100233

Sheskin, M., Bloom, P., & Wynn, K. (2014). *Anti-equality: Social comparison in young children*. *Cognition*, *130*(2), http://doi.org/10.1016/j.cognition.2013.10.008.

Silberberg, A., Crescimbene, L., Addessi, E., Anderson, J. R., & Visalberghi, E. (2009). Does inequity aversion depend on a frustration effect? A test with capuchin monkeys (Cebus apella). *Animal Cognition*, *12*(3), 505–509. http://doi.org/10.1007/s10071-009-0211-6

Silk, J. B. (2009). Chapter 18 – Social Preferences in Primates. In *Neuroeconomics* (pp. 269–284). http://doi.org/10.1016/B978-0-12-374176-9.00018-X

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

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

Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I Should but I Won't: Why Young Children Endorse Norms of Fair Sharing but Do Not Follow Them. *PLoS ONE*, *8*(3), e59510. http://doi.org/10.1371/journal.pone.0059510

Smith, P., & Silberberg, A. (2010). Rational maximizing by humans (Homo sapiens) in an ultimatum game. *Animal Cognition*, *13*(4), 671–677. http://doi.org/10.1007/s10071-010-0310-4

Smith, R. H., Eyre, H. L., Powell, C. A. J., & Kim, S. H. (2006). Relativistic origins of emotional reactions to events happening to others and to ourselves. *British Journal of Social Psychology*, *45*(2), 357–371. http://doi.org/10.1348/014466605X40987

Smith, R. H., Powell, C. A. J., Combs, D. J. Y., & Schurtz, D. R. (2009). Exploring the When and Why of Schadenfreude. *Social and Personality Psychology Compass*, *34*(10), 530–546. http://doi.org/10.1111/j.1751-9004.2009.00181.x

Smith, R. H., Turner, T. J., Garonzik, R., Leach, C. W., Urch-Druskat, V., & Weston, C. M. (1996). Envy and Schadenfreude. *Personality and Social Psychology Bulletin*, *22*(2), 158–168.

Smith, R. H., Turner, T. J., Garonzik, R., Leach, C. W., Urch-Druskat, V., & Weston, C. M. (1996). Envy and Schadenfreude. *Personality and Social Psychology Bulletin*, *22*(2), 158–168. http://doi.org/10.1177/0146167296222005

Socarides, C. W. (1977). *The World of emotions : clinical studies of affects and their expression*. New York: International Universities Press.

Solnick, S. (2001). Gender differences in the ultimatum game. *Economic Inquiry*, *39*(2), 189–200. http://doi.org/10.1111/j.1465-7295.2001.tb00060.x

Solomon, R. C. (1994). Sympathy and Vengeance: The Role of the Emotions in Justice. In S. van Goozen, N. van de Poll, & J. Sergeant (Eds.), *Emotions: Essays in Emotion Theory* (pp. 291–311). Hillsdale, NJ: Lawrence Erlbaum.

Steinbeis, N., & Singer, T. (2013). The effects of social comparison on social emotions and behavior during childhood: The ontogeny of envy and Schadenfreude predicts developmental changes in equity-related decisions. *Journal of Experimental Child Psychology*, *115*(1), 198–209. http://doi.org/10.1016/j.jecp.2012.11.009

Stuckless, N., & Goranson, R. (1992). The vengeance scale: Development of a measure of attitudes toward revenge. *Journal of Social Behaviour and Personality*, *7*(1), 25–42. http://doi.org/0886-1641

Sueiro, D. (1974). *La pena de muerte, ceremonial, historia, procedimientos*. Madrid: Alianza.

Summerfield, D. (2002). Effects of War: Moral Knowledge, Revenge, Reconciliation, and Medicalised Concepts of "Recovery." *British Medical Journal*, *325*, 1105–1107. http://doi.org/10.1136/bmj.325.7372.1105

Sutter, M. (2007). Outcomes versus intentions: On the nature of fair behavior and its development with age. *Journal of Economic Psychology*, *28*(1), 69–78. http://doi.org/10.1016/j.joep.2006.09.001

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., & Okubo, Y. (2009). When Your Gain Is My Pain and Your Pain Is My Gain: Neural Correlates of Envy and Schadenfreude. *Science*, *323*(5916).

Taylor, B. (2008). *The Book of Getting Even: a Novel*. Hanover, New Hampshire: Steerforth Press.

Tisak, M. S. (1993). Preschool children's judgments of moral and personal events involving physical harm and property damage. *Merrill-Palmer Quarterly*, *39*(3), 375–390.

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two Key Steps in the Evolution of Human Cooperation. *Current Anthropology*, *53*(6), 673–692. http://doi.org/10.1086/668207

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57. http://doi.org/10.1086/406755

Tuchman, B. W. (2011). *A Distant Mirror: The Calamitous 14th Century*. New York: Random House Publishing Group.

Tunick, M. (1992). *Punishment: Theory and Practice*. Berkeley & L.A.: University of California Press.

Ugidos, G. (2015). *Grandes venganzas de la historia. El verdadero conde de Monte Cristo y otros ajustes de cuentas memorables.* Madrid: La Esfera de los Libros.

Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young Children Selectively Avoid Helping People With Harmful Intentions. *Child Development*, *81*(6), 1661–1669. http://doi.org/10.1111/j.1467-8624.2010.01500.x

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology*, *29*(1), 124–130. http://doi.org/10.1348/026151010X532888

Valiente, C., Eisenberg, N., Fabes, R. A., Shepard, S. A., Cumberland, A., & Losoya, S. H. (2004). Prediction of Children's Empathy-Related Responding From Their Effortful Control and Parents' Expressivity. *Developmental Psychology*, *40*(6), 911–926. http://doi.org/10.1037/0012-1649.40.6.911

Vidmar, N. (2000). Retribution and revenge. In J. Sanders & V. L. Hamilton (Eds.), *Handbook of justice research in law* (pp. 31–63). New York: Kluwer Academic Publishers-Plenum Publishers.

Vogelsang, M., Jensen, K., Kirschner, S., Tennie, C., & Tomasello, M. (2014). Preschoolers are sensitive to free riding in a public goods game. *Frontiers in Psychology*, *5*, 729. http://doi.org/10.3389/fpsyg.2014.00729

von Rohr, C. R., Koski, S. E., Burkart, J. M., Caws, C., Fraser, O. N., Ziltener, A., … Marlowe, F. (2012). Impartial Third-Party Interventions in Captive Chimpanzees: A Reflection of Community Concern. *PLoS ONE*, *7*(3), e32494. http://doi.org/10.1371/journal.pone.0032494

von Uexküll, J. (1920). *Biologische Briefe an eine Dame*. (Paetel, Ed.). Retrieved from https://books.google.es/books/about/Biologische_Briefe_an_eine_Dame.html?id=n3IgnQEACAAJ&redir_esc=y

Warren, C., & Mcgraw, A. P. (2015). Journal of Personality and Social Psychology Differentiating What Is Humorous From What Is Not Differentiating What Is Humorous From What Is Not. *Journal of Personality and Social Psychology*. http://doi.org/10.1037/pspi0000041

West, S. A., & Gardner, A. (2010). Altruism, Spite, and Greenbeards. *Science*, *327*(5971).

Whiten, A., Custance, D. M., Gomez, J. C., Teixidor, P., & Bard, K. A. (1996). Imitative learning of artificial fruit processing in children (Homo sapiens) and chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *110*(1), 3–14. http://doi.org/10.1037/0735-7036.110.1.3

Wiessner, P. (2005). Norm enforcement among the Ju/'hoansi Bushmen. *Human Nature*, *16*(2), 115–145. http://doi.org/10.1007/s12110-005-1000-9

Wittig, M., Jensen, K., & Tomasello, M. (2013). Five-year-olds understand fair as equal in a mini-ultimatum game. *Journal of Experimental Child Psychology*, *116*(2), 324–337. http://doi.org/10.1016/j.jecp.2013.06.004

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, Act, and Outcome in Behavioral Prediction and Moral Judgment. *Child Development*, *67*(5), 2478–2492. http://doi.org/10.1111/j.1467-8624.1996.tb01869.x